Оглавление

Глава 1. Моделирование систем массового обслуживания 1

I. Основы теории массового обслуживания 1

II. Основные понятия СМО 4

III. Классификация СМО 7

IV. Показатели эффективности СМО 8

V. Потоки событий (заявок) 9

V.1. Пуассоновский поток 10

V.2. Пример АЗС 15

V.3 Потоки Эрланга 23

Глава 1. Моделирование систем массового обслуживания

I. Основы теории массового обслуживания

Большой класс систем, которые сложно изучать аналитическими способами, но которые хорошо изучаются методами статистического моделирования, сводится к СМО (системам массового обслуживания).

ТМО рассматривает математические модели систем, призванных обслуживать случайно возникающие требования.

Основоположником теории МО считается датский ученый А.К. Эрланг (1878 – 1929). Являясь сотрудником копенгагенской телефонной компании, он опубликовал в 1909 году работу «Теория вероятностей и телефонные переговоры», в которой решил ряд задач по теории СМО с отказами.

Значительный вклад в создание и разработку общей теории МО внес выдающийся советский математик А.Я. Хинчин (1884-1959), который предложил сам термин «теория массового обслуживания». В зарубежной литературе чаще используется название «теория очередей».

ТМО – это раздел теории вероятностей, целью исследования которого является рациональный выбор структуры системы обслуживания и процесса обслуживания на основе изучения:

а) потоков требований на обслуживание, поступающих в систему и выходящих из неё;
б) длительности ожидания;
в) длины очередей.

ТМО нашла применение при решении множества различных задач, что является следствием существования многих типов систем обслуживания случайно приходящих требований.

В качестве примеров СМО в финансово-экономической среде можно привести системы, представляющие собой:

1) банки;
2) страховые организации;
3) налоговые инспекции;
4) аудиторские службы.

В сфере производства и обслуживания:

1) различные системы связи (в том числе телефонные станции);
2) погрузочно-разгрузочные комплексы (порты, товарные станции);
3) АЗС;
4) магазины;
5) парикмахерские;
6) билетные кассы;
7) пункты обмена валюты;
8) ремонтные мастерские;
9) больницы (прием пациентов в клиниках);
10) регулирование переключения светофоров.

Такие системы как компьютерные сети, системы сбора, хранения и обработки int; транспортные системы; автоматизированные производственные участки; и в военной области: системы противовоздушной или противоракетной обороны также могут рассматриваться как своеобразные СМО. Несмотря на кажущееся различие всех перечисленных выше задач, они имеют дело со следующей ситуацией: «требование» поступает в «счетчик», фиксирующий его приход, и требует обслуживания. Если обслуживающий прибор занят другим требованием, вновь прибывшее требование должно или ждать его освобождения (до полного освобождения, либо некоторое ограниченное время), или уходит, т.е. «теряется». В это время на обслуживание могут поступать другие требования. Если требования поступают в то время, когда прибор занят, они образуют очередь или линию ожидания. Здесь мы употребляем термины «требования», «счетчик» и «прибор» в их самом общем смысле. В каждом частном случае эти термины можно определить более точно. Например, в задачах телефонии «счетчиком» называется телефонная станция, «требование» – поступающий вызов, а «обслуживающий прибор» – телефонная линия или канал.

Для того чтобы описать СМО, необходимо задать следующие ее компоненты:

1) входной поток;
2) дисциплину очереди;
3) механизм обслуживания.

Обобщенная схема СМО приведена на рис. 1:

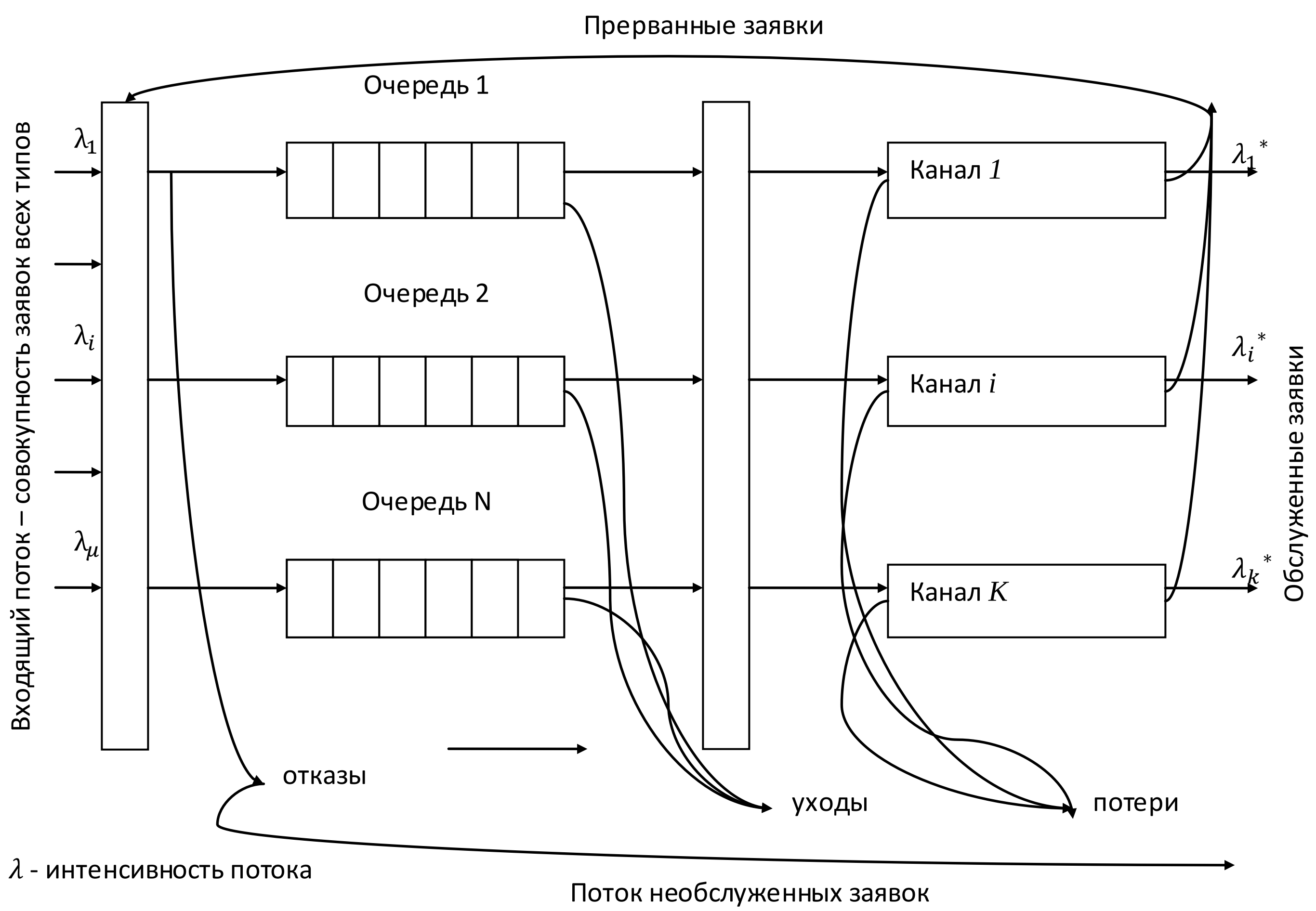


рис.1, Схема СМО

Пример. СМО: автобусный маршрут. Заявки: пассажиры. Канал: автобусы.

Таким образом, во всякой СМО можно выделить следующие основные элементы:

1) входящий поток заявок;
2) очередь;
3) каналы обслуживания;
4) выходящий поток обслуженных заявок.

Заявки могут приходить неравномерно, каналы могут обслуживать разные заявки в разное время и т.д., количество заявок всегда весьма велико.

Все это делает такие СМО сложными для изучения и управления, и проследить все причинно-следственные связи в них не представляется возможным. Поэтому принято представление о том, что обслуживание в сложных системах носит случайный характер.

Подход к изучению СМО с помощью методов компьютерного моделирования заключается в том, что:

1) с помощью генератора случайных чисел разыгрываются случайные числа, которые имитируют случайные моменты появления заявок и время их обслуживания в каналах.
2) с использованием типовых элементов (канал, источник заявок, очередь, заявка, дисциплина обслуживания, стек, кольцо и т.д.) имитируется деятельность СМО (продвижение заявок по каналам; покидания канал заявками; потери заявок; обработка заявок) и анализируются различные показатели (вероятность обслуживания заявки системой; среднее количество занятых каналов и т.д.)

По результатам этого анализа специалист оценивает ресурсы производительности и эффективности проектируемых им систем, скрытые в оптимизации параметров, структур и дисциплин обслуживания. Моделирование помогает выявить эти скрытые резервы.

II. Основные понятия СМО

Источник заявок порождает заявки в случайные моменты времени, согласно заданному пользователем законом распределения.

Заявки (или клиенты) порождаются источником заявок, входят в систему, проходят через ее элементы (обслуживаются), покидают ее обслуженными или необслуженными.

Бывают нетерпеливые заявки – такие, которым надоело ждать или находиться в системе и которые покидают по собственной воле СМО.

Заявки образуют потоки – поток заявок на входе системы, поток обслуженных заявок, поток необслуженных заявок. Поток характеризуется количеством заявок определенного сорта, наблюдаемым в некотором месте СМО за единицу времени (час, сутки, месяц), т.е. поток – это величина статистическая.

Каналы – то, что обслуживает. Бывают горячие – те, которые начинают обслуживать заявку в момент ее поступления в канал, и холодные – каналу для начала обслуживания требуется время на подготовку.

Очереди характеризуются:

1) правилами пребывания в очереди (т.е. дисциплиной постановки в очередь и выбора из неё);
2) количеством мест в очереди (сколько максимально может находиться одновременно заявок в ней);
3) структурой очереди (связь между местами в очереди).

Бывают ограниченные и неограниченные очереди.

К важнейшим дисциплинам обслуживания относятся:

1) FIFO: если заявка первой пришла в очередь, то она первой и уйдет на обслуживание;
2) LIFO: если заявка пришла в очередь последней, то она первой обслуживается;
3) SF (short forward): в первую очередь обслуживаются те заявки из очереди, которые имеют меньшее время обслуживания;
4) случайный выбор из очереди;
5) выбор из очереди по параметрам: сначала получают помощь больные с температурой, а потом «здоровые» больные.

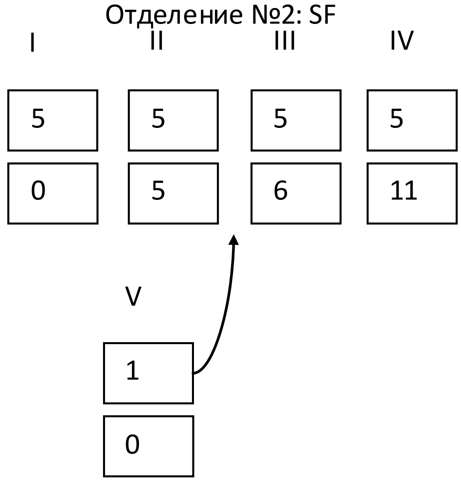
- **Пример**
  Как правильный выбор той или иной дисциплины обслуживания позволяет получить ощутимую экономию во времени.

Пусть имеется два отделения почты.

Отделение №1: FIFO

$$t_{\text{среднее ожидания}} = \frac{0+5+\cdots+20}{5} = 10\ \text{мин}$$

Отделение №2: SF



$$t_{\text{среднее ожидания}} = \frac{0+5+0+6+11}{5}$$
$$= 4{,}4\ \text{мин}$$

III. Классификация СМО

1. По числу каналов:
   1.1. одноканальные;
   1.2. многоканальные;
      1.2.1. однородные каналы;
      1.2.2. разнородные каналы (отличаются длительностью обслуживания одной заявки).
2. По дисциплинам обслуживания:
   2.1. СМО с отказами, в которых заявка, поступившая на вход СМО в момент, когда все каналы заняты, получает «отказ» и покидает СМО («пропадает»). Чтобы эта заявка все же была обслужена, она должна снова поступить на вход СМО и рассматриваться при этом как заявка, поступившая впервые. Примером СМО с отказами может служить работа АТС: если набранный телефонный номер (заявка, поступившая на вход) занят, то заявка получает отказ, и, чтобы дозвониться по этому номеру, следует его набрать еще раз (заявка поступает на вход как новая);
   2.2. СМО с ожиданием (неограниченным ожиданием или очередью). В таких системах заявка, поступившая в момент занятости всех каналов, становится в очередь и ожидает освобождения канала, который примет ее к обслуживанию. Каждая заявка, поступившая на вход, в конце концов, будет обслужена. Такие СМО часто встречаются в торговле, в сфере бытового и медицинского обслуживания, на предприятиях (например, обслуживание станков бригадой наладчиков);
   2.3. СМО смешанного типа (с ограниченным ожиданием). Это такие системы, в которых на пребывание заявки в очереди накладываются некоторые ограничения. Эти ограничения могут накладываться на длину очереди, т.е. максимально возможное число заявок, которые одновременно могут находиться в очереди. В качестве примера такой системы можно привести мастерскую по ремонту автомобилей, имеющую ограниченную по размерам стоянку для неисправных машин, ожидающих ремонта. Ограничения ожидания могут касаться времени пребывания заявки в очереди, по истечению которого она выходит из очереди и покидает систему, либо касаться общего времени пребывания заявки в СМО (т.е. суммарного времени пребывания заявки в очереди и под обслуживанием);
3. По ограничению потока заявок:
   3.1. замкнутые, если поток заявок ограничен и заявки, покинувшие систему, могут в нее возвращаться;
   3.2. открытые;

4. По количеству этапов обслуживания:
   4.1. однофазные, если каналы СМО однородны, т.е. выполняют одну и ту же операцию обслуживания;
   4.2. многофазные, если каналы обслуживания расположены последовательно и они неоднородны, так как выполняют различные операции обслуживания (т.е. обслуживание состоит из нескольких последовательных этапов или фаз).

IV. Показатели эффективности СМО

При анализе результатов моделирования СМО важно указать интересы и степень их выполнения. Различают интересы клиента (заявки) и интересы владельца системы. Заметим, что эти интересы не всегда совпадают.

В качестве характеристик эффективности функционирования СМО можно выбрать три основные группы (обычно средних) показателей:

1. Показатели эффективности использования СМО:
   1.1. Абсолютная пропускная способность СМО – среднее число заявок, которое сможет обслужить СМО в единицу времени.
   1.2. Относительная пропускная способность СМО – отношение среднего числа заявок, обслуживаемых СМО в единицу времени, к среднему числу поступивших за это же время заявок.
   1.3. Средняя продолжительность периода занятости СМО.
   1.4. Коэффициент использования СМО – средняя доля времени, в течение которого СМО занята обслуживанием заявок, и т.п.
2. Показатели качества обслуживания заявок:
   2.1. Среднее время ожидания заявки в очереди.
   2.2. Среднее время пребывания заявки в СМО.
   2.3. Вероятность отказа заявке в обслуживании без ожидания.
   2.4. Вероятность того, что вновь поступившая заявка немедленно будет принята к обслуживанию.
   2.5. Закон распределения времени ожидания заявки в очереди.
   2.6. Закон распределения времени пребывания заявки в СМО.
   2.7. Среднее число заявок, находящихся в очереди.
   2.8. Среднее число заявок, находящихся в СМО, и т.п.
3. Показатели эффективности функционирования пары «СМО – клиент», где под «клиентом» понимают всю совокупность заявок или некий их источник. К числу таких показателей относится, например, средний доход, приносимый СМО в единицу времени, и т.п.

Судить о качестве полученной системы нужно по совокупности этих показателей. При анализе результатов моделирования (т.е. показателей) важно обращать внимание на интересы клиента и интересы владельца системы, т.е. надо минимизировать или максимизировать тот или иной показатель, а также степень их выполнения. Показатели оптимизируются за счет изменения параметров СМО. К ним относятся:

- интенсивность потока заявок;
- интенсивность потока обслуживания;
- среднее время, в течение которого заявка готова ожидать обслуживания в очереди;
- количество каналов обслуживания;
- дисциплина обслуживания каналов и т.д.

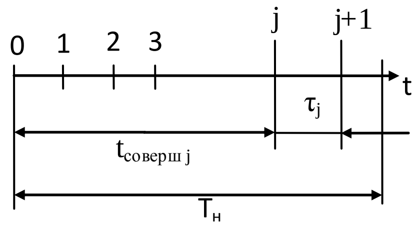
Заметим, что владелец системы может влиять не на все параметры (например, не может повлиять на интенсивность потока заявок).

V. Потоки событий (заявок)

Под потоком событий понимается последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени (например, поток вызовов на телефонной станции; поток покупателей; поток заказных писем, поступающих в почтовое отделение, и т.п.). Поток характеризуется интенсивностью λ – частотой появления событий или средним числом событий, поступающих в СМО в единицу времени.

Интенсивность потока в некотором смысле является математическим ожиданием количества событий в единицу времени. Но реально может оказаться так, что в один час появится 4 события, в другой – 6, в среднем получается 5 событий в час; поэтому вторая величина, характеризующая насколько велик разброс событий относительно математического ожидания – дисперсия. Собственно именно эта величина определяет случайность появления события, слабую предсказуемость момента его появления.

Таким образом, поток событий – это последовательность однородных событий, поступающих одно за другим в случайные промежутки времени.



τ_j – интервал между событиями (случайная величина),

$t_{\text{соверш } j}$ – момент совершения j-ого события (отсчитывается от t=0).

$T_н$ – время наблюдения

Интенсивность потока λ – это среднее число событий в единицу времени.

$\lambda = \frac{N}{T_н}$, где N – число событий, произошедших за время наблюдения $T_н$.

Если интервал между событиями равен константе или определен какой-либо формулой $\tau_j = f(\tau_{j-1})$, то поток называется регулярным (поток изделий на конвейере сборочного цеха – с постоянной скоростью движения). Такой поток встречается довольно редко в реальных системах. Иначе – случайный поток заявок:

1) Стационарные (вероятностные характеристики не зависят от времени, $\lambda(t) = \lambda$);
2) Поток без последствий (если для любых двух непересекающихся участков на временной оси τ_1 и τ_2 число событий, попадающих на один из них, не зависит от числа событий, попавших на другой, т.е. вероятность появления случайного события не зависит от момента совершения предыдущих событий).
3) Ординарные (вероятность одновременного появления двух и более событий равна 0, т.е. события появляются по очереди, а не группами).

Поток событий называется простейшим (или стационарным Пуассоновским), если одновременно он стационарен, ординарен и не имеет последствий (простейший поток в том смысле, что имеет наиболее простое математическое описание. Между прочим, самый простой, на первый взгляд, регулярный поток не является «простейшим», т.к. обладает последствием: моменты появления событий в таком потоке связаны жесткой функциональной зависимостью. Без специальных усилий по поддержанию его регулярности такой поток обычно не создаётся).

V.1. Пуассоновский поток

Напомним, что в общем случае Пуассоновский поток является ординарным, и потоком без последствий.

Случайная величина, которая характеризует количество заявок в потоке, поступающих в СМО на временном интервале $[t_0; t_0+\tau]$, распределена по закону Пуассона (вероятность появления m заявок за время τ):

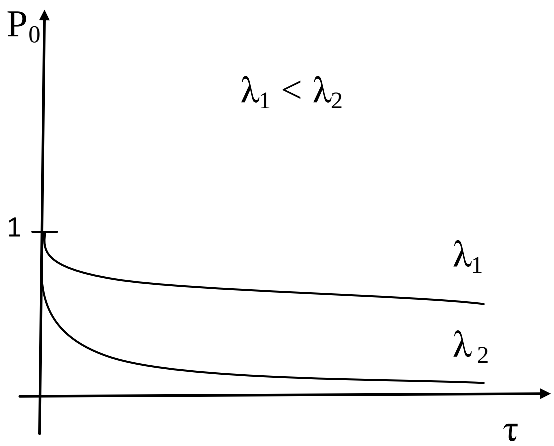
$P_m = \frac{a^m e^{-a}}{m!}$, где m – количество заявок, a – параметр Пуассона, $a = \tau\lambda$, λ – интенсивность или плотность потока (среднее число заявок в единицу времени).

$$\begin{cases} M = \lambda - \text{математическое ожидание} \\ \sigma^2 = \lambda^2 - \text{дисперсия} \end{cases}$$

В общем случае параметр Пуассона рассчитывается как: $a = \int_{t_0}^{t_0+\tau} \lambda(t)dt$, т.е.

a – математическое ожидание числа событий (среднее число событий в рассматриваемом интервале τ).

Если $\lambda(t)$ = const, то это стационарный поток Пуассона (простейший), и в этом случае $a=\lambda\tau$, иначе – нестационарный поток.



Вероятность непоявления события для простейшего потока в зависимости от времени наблюдения.

Вероятность непоявления событий (т.е. когда m = 0): $P_0 = e^{-a}$.

Если интенсивность потока велика, то вероятность **не**появления события быстро уменьшается (быстрее наступит событие).

Вероятность появления хотя бы одного события:

$P^* = 1 - P_0 = 1 - e^{-a}$

Очевидно, что при возрастании времени наблюдения вероятность появления хотя бы одного события стремится к единице: $P^* \to 1$ (событие рано или поздно произойдёт).

Таким образом, в случае простейшего потока ($a=\lambda\tau$) длительность временного интервала между двумя последовательными заявками имеет показательное распределение с параметром λ:

$F_\tau(t) = \begin{cases} 1 - e^{-\lambda t}, t \geq 0 \\ 0, t < 0 \end{cases}$, где λ – параметр масштаба, интенсивность случайной величины.

$$\begin{cases} M = 1/\lambda - \text{математическое ожидание} \\ \sigma^2 = 1/\lambda^2 - \text{дисперсия} \end{cases}$$

Это распределение – единственное непрерывное распределение, обладающее свойством отсутствия последствий, т.е. $\forall x_0 > 0$ и $x > 0$: $P(X - x_0 < x | X \geq x_0) = P(X)$. Это свойство называется Марковским свойством.

- **Пример**

Чтобы смоделировать Пуассоновский простейший поток, надо смоделировать временной интервал между заявками (для этого воспользуемся методом обратной функции):

$$\alpha = 1 - e^{-\lambda\tau}, \alpha \sim Rav[0,1].$$

$$\boldsymbol{\tau = -\frac{1}{\lambda}\ln(1-\alpha) \approx -\frac{1}{\lambda}\ln\alpha}. \qquad (1)$$

Рассмотрим поток покупателей интернет-магазина, приходящих в среднем 6 человек в сутки. Необходимо промоделировать процесс в течение 100 часов.

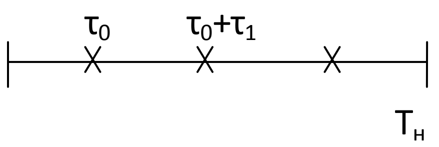
Значит, интенсивность потока: $\lambda = \frac{6}{24} = \frac{1}{4}\left(\frac{\text{ед}}{\text{час}}\right)$.

Время наблюдения: $T_н = 100$(час).

Математическое ожидание (среднее время) длительности между двумя покупателями: $m = \frac{1}{\lambda} = 4$. (т.е. в среднем приходит один покупатель за 4 часа).

$\sigma = 4$, $t = 0$, τ_0 по формуле (1), $t = \tau_0$, τ_1 по формуле (1),

$t = \tau_0 + \tau_1$, ..., while $t \leq T_н$.



Количество смоделированных τ_i равно количеству покупателей за $T_н$.

В среднем это количество будет равно

$\frac{6*100}{24} = 25$ человек за 30100 часов.

- **Пример**

Приведем пример моделирования неординарных событий (когда в один момент может появиться несколько событий). Т.е. необходимо моделировать и число событий. Например, вагоны на ж/д станцию пребывают в составе поезда в случайные моменты времени: ординарный поток поездов.

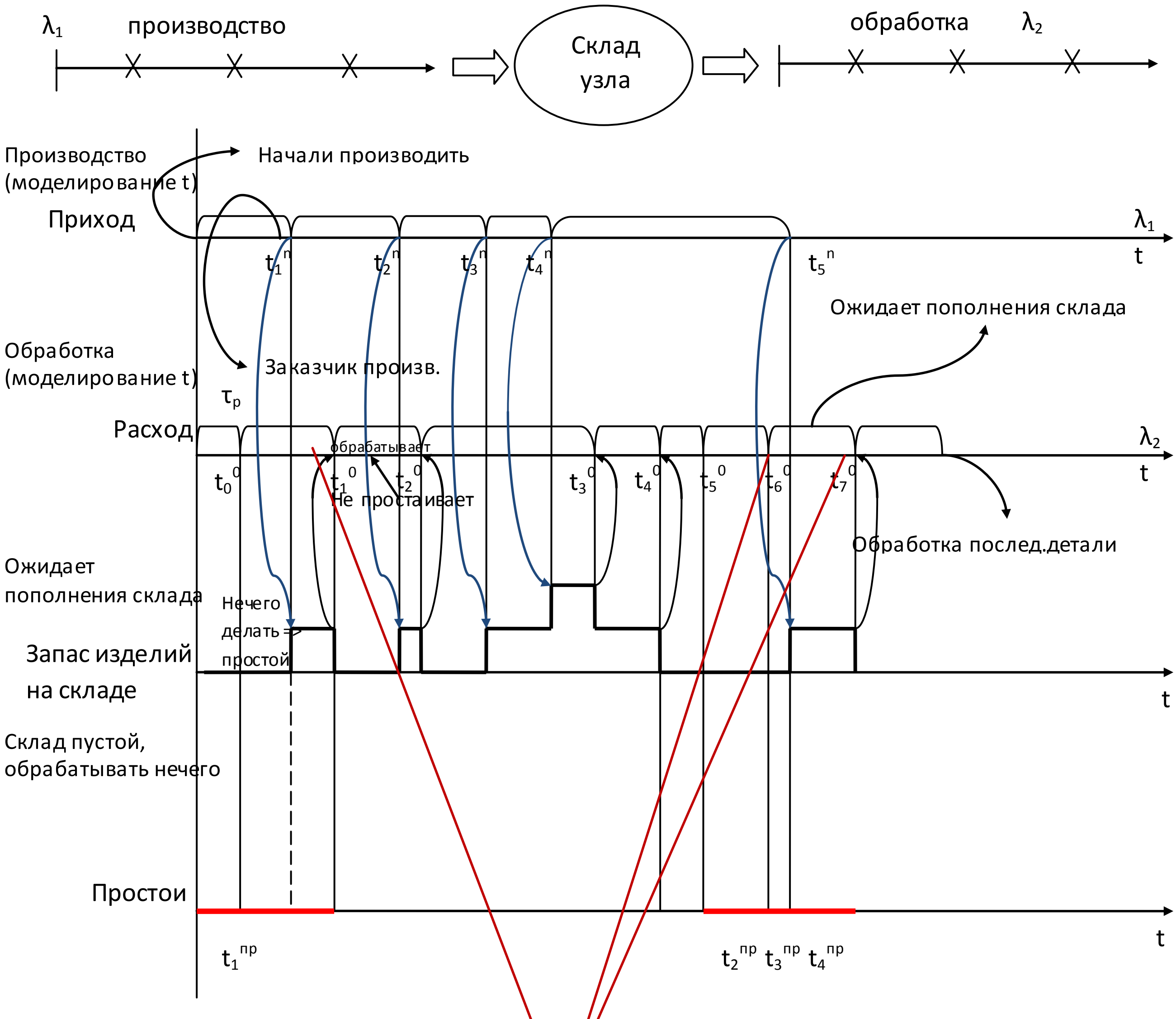
При этом в составе поезда может быть разное (случайное) количество вагонов. В этом случае о потоке вагонов говорят как о потоке неординарных событий.

Пусть в среднем в 68 случаях за 100 приходит от 6 до 14 вагонов в составе поезда и их число распределено по нормальному закону, т.е. $M = \frac{(6+14)}{2} = 10, \sigma = 4$. Т.е. после того, как мы смоделировали τ_i, надо смоделировать число вагонов.

$$z = \sqrt{-2\ln(\alpha_1)} \begin{cases} \sin \\ \cos \end{cases} (2\pi\alpha_2), \alpha_1, \alpha_2 \sim \mathrm{Rav}[0,1]. \qquad x = (m + \sigma z).$$

- **Пример**

Каково среднее время суточного простоя оборудования технологического узла, если узел обрабатывает каждое изделие случайное время, заданное интенсивностью потока случайных событий λ_2? При этом экспериментально установлено, что производят изделия на обработку тоже в случайные моменты времени, заданные потоком λ_1 партиями. До начала моделирования T=0 изделий на складе не было. Необходимо промоделировать этот процесс в течение $T_н$=100.



$$\overline{\mathrm{T}}_{\text{простое}} = 24(t_1^{\text{пр}} + \cdots)/\mathrm{T}_{\text{н}}$$

Закончилось t, отведенное на обработку «виртуальной» детали, которой на самом деле не было.

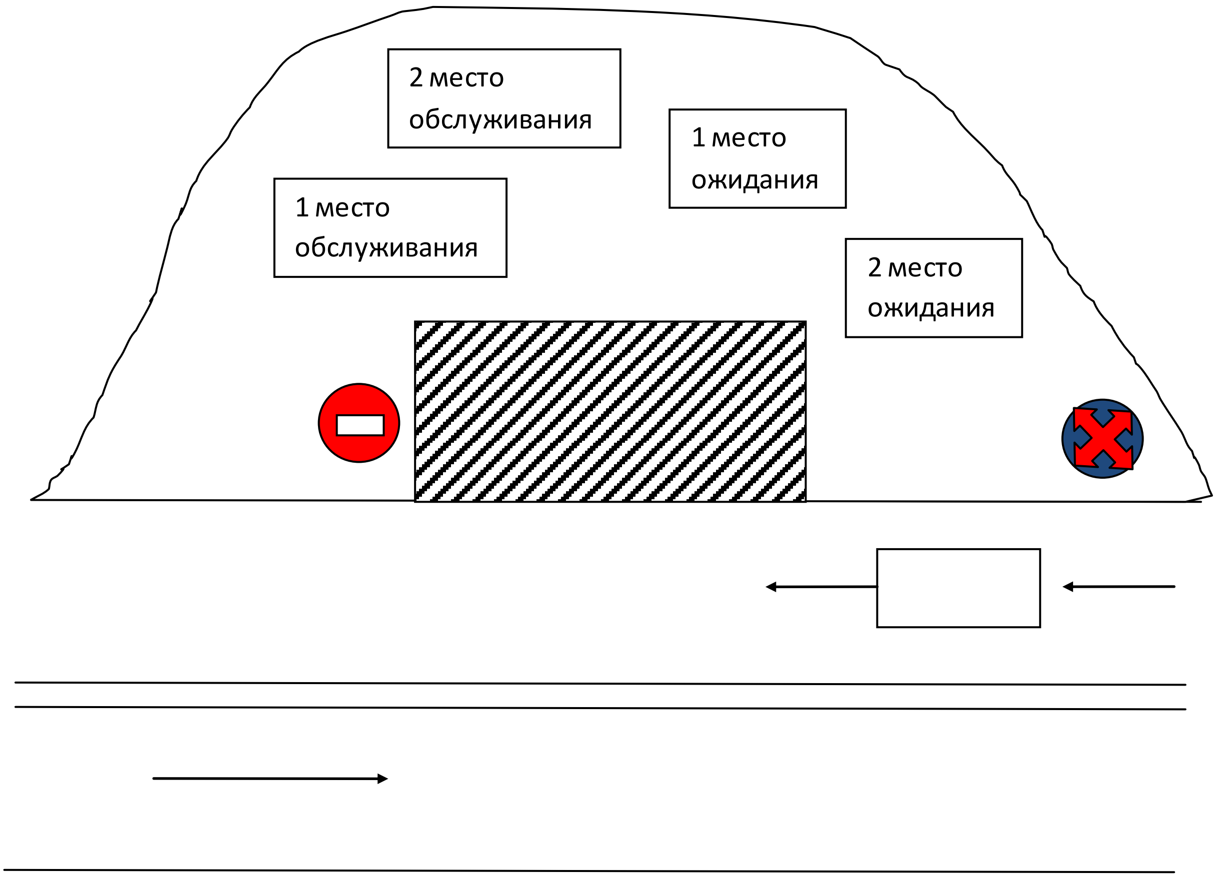
- **Замечание**

Если поток нестационарный, т.е. λ=λ(t), то есть зависит от t, например, поток покупателей в магазине: $\lambda(t) = \begin{cases} 7^{00} - 9^{00} & 55\frac{\text{чел}}{\text{час}} \\ 9^{00} - 12^{30} & 45\frac{\text{чел}}{\text{час}} \\ 0^{00} - 3^{00} & 1\frac{\text{чел}}{\text{час}} \end{cases}$ Может быть задано в виде графиков, таблиц, формул.

то это необходимо учитывать при моделировании интервалов возникновения событий.

V.2. Пример АЗС

Рассмотрим метод моделирования СМО на примере АЗС, а так же план её исследования.

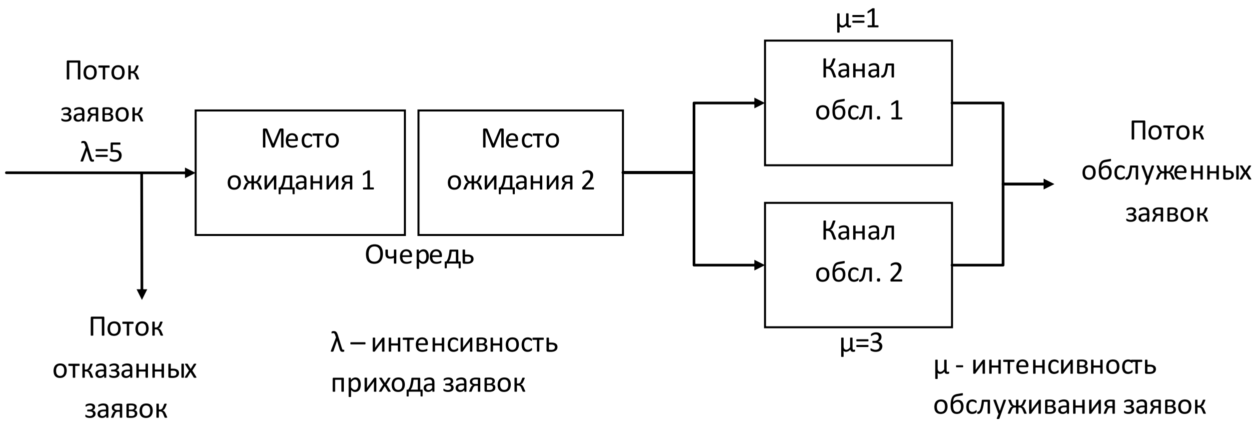


Водитель, проезжая по дороге, может захотеть заправить свой автомобиль. Хотят этого не все водители. Из всего потока машин на заправку в среднем заезжает 5 машин в час.

Имеем 2 одинаковые колонки, статистическая производительность каждой из которых известна. Первая колонка обслуживает в среднем одну машину в час, вторая – три в час.

Если колонки заняты, то на месте ожидания могут ожидать обслуживания другие машины. Если появилась третья машина, а все места в очереди заняты, то ей отказывают в обслуживании, т.к. стоять на дороге запрещено законом. Такая машина уезжает прочь из системы навсегда и как потенциальный клиент является потерянным для владельца АЗС. (АЗС будем рассматривать без кассы))

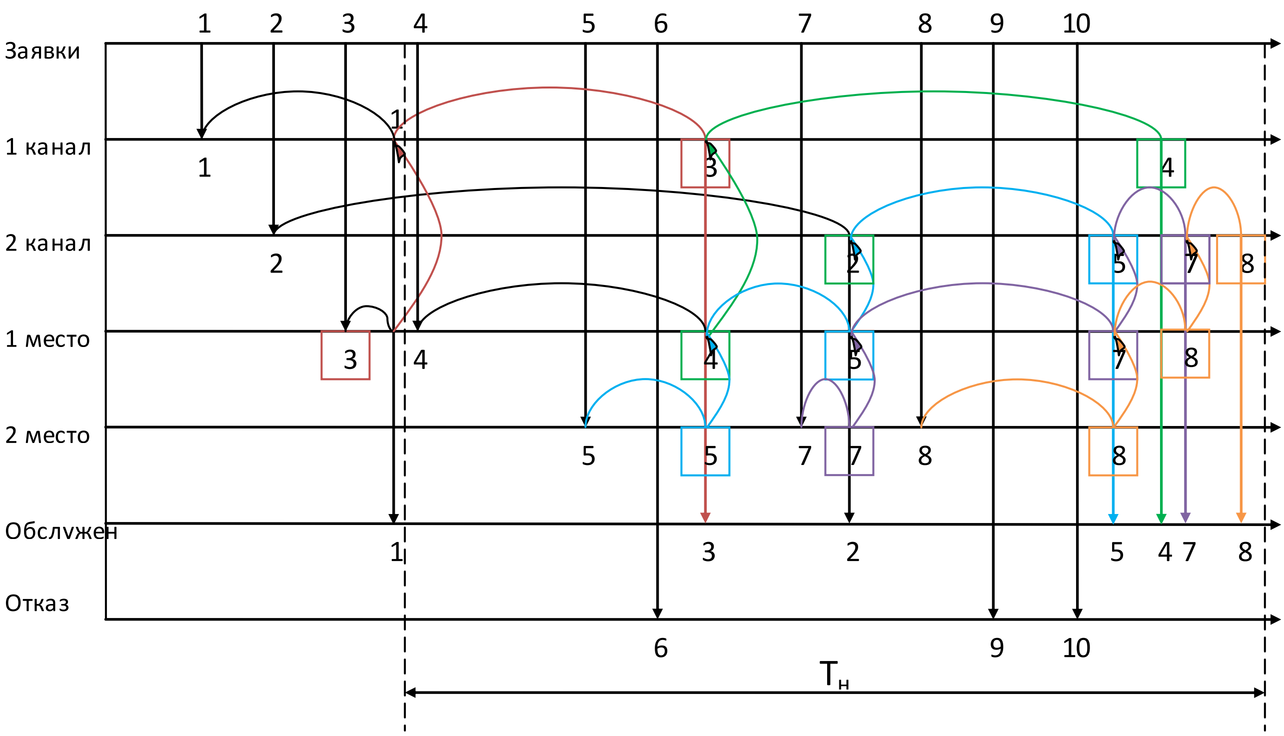
Пути движения потоков заявок по СМО можно изобразить в виде эквивалентной схемы, а добавив значения и обозначения характеристики каждого элемента СМО, получаем окончательную схему:



На этой схеме приведен принцип последовательной проводки заявок.

Для наглядности простроим временную диаграмму работы СМО, отражая на каждой линейке (ось времени t) состояние отдельного элемента системы.

Для генерации времени прихода заявок используем формулу (1), вычисляя интервал времени τ между моментами прихода двух случайных событий.



Временная диаграмма работы СМО

После того, как первая заявка прошла весь путь, можно согласно принципу последовательной проводки заявок также проимитировать путь второй заявки.

Если в какой-то момент окажется, что оба канала заняты, то следует установить заявку в очередь (заявка №3). Заметим, что по условию задачи в очереди в отличие от каналов заявки находятся не случайное время, а ожидают, когда освободится какой-то из каналов. После освобождения канала заявка поднимается на линейку соответствующего канала и там организуется ее обслуживание. Если все места в очереди в момент, когда придет очередная заявка, будут заняты, то заявку следует отправить на линейку «Отказанные» (заявка №6).

Процедуру имитации обслуживания заявок продолжают некоторое время наблюдения $Т_н$. ***Чем больше это время, тем точнее в дальнейшем будут результаты моделирования.*** Реально для простых систем выбирают $Т_н$ равное 50-100 или более часов, хотя иногда лучше мерить эту величину количеством рассмотренных заявок.

Далее рассмотрим анализ временной диаграммы.

Сначала надо дождаться установившегося режима. Откидываем первые 4 заявки как нехарактерные, протекающие во времени процесса установления работы системы. Измеряем время наблюдения. Допустим, $Т_н$=5 часов. Подсчитываем из диаграммы количество обслуженных заявок ($N_{обслуж}$), время простоя ($t_{простоя}$) и другие величины. В результате можем вычислить показатели, характеризующие качество работы СМО.

1. Вероятность обслуживания:
   $$P_{обслуж} = \frac{N_{обслуж}}{N} = \frac{6}{7} = 0{,}857,$$
   где $N_{обслуж}$ показывает, сколько заявок удалось обслужить за $Т_н$ (заявки №2,3,4,5,7,8), N – сколько всего заявок пришло за $Т_н$ (заявки № 4-№ 10).
2. Пропускная способность системы:
   $$A = \frac{N_{обслуж}}{Т_н} = \frac{7}{5} = 1{,}4\left(\frac{шт}{час}\right)$$
3. Вероятность отказа:
   $$P_{отказ} = \frac{N_{отказ}}{N} = \frac{3}{7} = 0{,}43.$$
   При этом $P_{обслуж}+P_{отказ} \neq 1$, т.к. $Т_н$ мало => погрешность.
4. Вероятность занятости одного канала:
   $$P_1 = \frac{Т_{зан}}{Т_н},$$
   где $Т_{зан}$ – время занятости только одного (первого или второго) канала.
5. Вероятность занятости двух каналов:
   $$P_2 = \frac{Т_{зан\ 2х}}{Т_н},$$

где $T_{зан2х}$ – время занятости одновременно всех каналов (первого и второго).

6. Среднее количество занятых каналов:

$$N_{с.к.} = 0 * P_0 + 1 * P_1 + 2 * P_2 = 0{,}01 + 2 * 0{,}99 = 1{,}99,$$

где P_1 – вероятность занятости одного канала, P_2 – вероятность занятости двух каналов. Эта цифра говорит о том, что из возможных двух каналов в среднем загружено 1,99 канала. Это высокий показатель нагрузки, т.е. $\frac{1{,}99}{2} * 100\% = 99{,}5\%$, система хорошо использует ресурс.

7. Вероятность простоя хотя бы одного канала:

$$P_1^* = \frac{T_{простоя1}}{T_н}.$$

8. Вероятность простоя двух каналов одновременно:

$$P_2^* = \frac{T_{простоя2}}{T_н} = 0.$$

9. Вероятность простоя всей системы:

$$P_c^* = \frac{T_{простоя\ сист}}{T_н} = 0.$$

10. Вероятность того, что в очереди будет одна заявка:

$$P_{1заяв} = \frac{T_{1заяв}}{T_н}.$$

11. Вероятность того, что в очереди будут стоять одновременно две заявки:

$$P_{2заяв} = \frac{T_{2заяв}}{T_н}.$$

12. Среднее количество заявок в очереди:

$$N_{с.з.} = 0 * P_{0заяв} + 1 * P_{1заяв} + 2 * P_{2заяв}.$$

13. Среднее время ожидания заявки в очереди:

$$T_{ср.ож.} = \frac{\sum_{i=1}^{N} T_{ож i}}{N},$$

т.е. надо сложить все временные интервалы, в течение которых какая-либо заявка находилась в очереди, и разделить на количество заявок.

14. Среднее время обслуживания заявки:

$$T_{ср.обслуж} = \frac{\sum_{i=1}^{N} T_{обслуж i}}{N},$$

т.е. сложить все интервалы времени, в течение которых какая-либо заявка находилась на обслуживании, и разделить на количество заявок.

15. Среднее время нахождения заявки в системе: $T_{ср.сист} = T_{ср.ож} + T_{ср.обслуж}$.

16. Среднее количество заявок в системе:

$$N_{ср} = \frac{\sum_{i=1}^{K}(N_{2i} + N_{3i} + N_{4i} + N_{5i})}{K}.$$

Разобьем $T_н$ на интервалы по 10 минут. Получится на 5 часах K подинтервалов (K=30). В каждом подинтервале определим по временной диаграмме, сколько заявок находится в системе. Смотреть надо 2-5 линейки (каналы и очереди) – какие из них заняты в данный момент. Затем сумму К слагаемых усреднить.

Далее следует составить таблицу результатов и оценить значения каждого из них с точки зрения клиента и владельца СМО. В конце, учитывая сказанное в каждом пункте, следует сделать общий вывод.

Показатель	Формула	Значение	Интересы владельца СМО	Интересы клиента СМО
Вероятность обслуживания	$P_{\text{обслуж}} = \frac{N_{\text{обслуж}}}{N}$	0,714	Вероятность обслуживания мала, много клиентов уходят из системы неудовлетворенными, их деньги для владельца потеряны. Это «минус». Рекомендация: увеличить вероятность обслуживания.	Вероятность обслуживания мала, каждый третий клиент хочет, но не может обслужиться. Это «минус». Рекомендация: увеличить вероятность обслуживания.
…	…	…	…	…
Среднее количество заявок в очереди	$N_{\text{с.з.}} = 0 * P_{0\text{заяв}} + 1 * P_{1\text{заяв}} + 2 * P_{2\text{заяв}}.$	1,62	Очередь практически все время забита. Все места в очереди используются достаточно эффективно. Вложения на организацию очереди окупают затраты на нее. Это «плюс». Клиенты, которые долго стоят в очереди, могут уйти, не дождавшись обслуживания. Клиенты, простаивая, могут	Очередь практически все время забита. Клиенту приходится стоять в очереди, прежде чем он попадет на обслуживание. Клиент может не попасть даже в очередь. Это «минус». Рекомендация: увеличить число мест в очереди, увеличить пропускную способность.

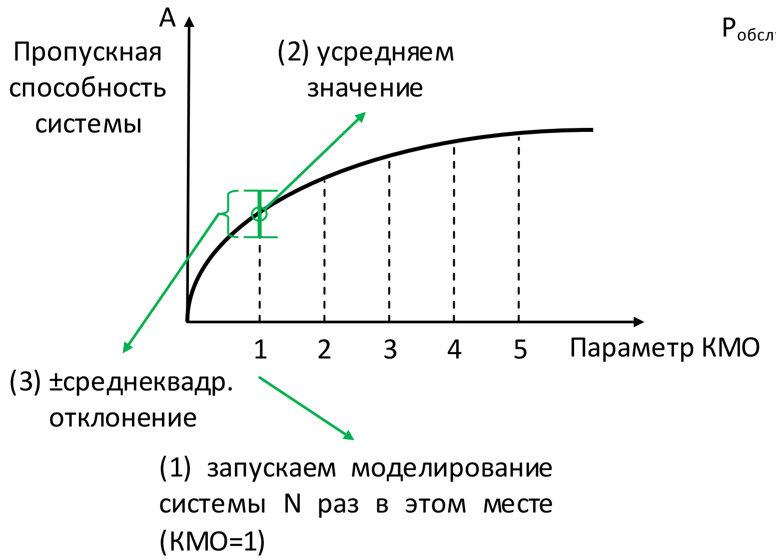
			нанести ущерб системе, ломать оборудование. Много отказов, потерянных клиентов. Это «минусы». Рекомендация: увеличить число мест в очереди, увеличить пропускную способность.	
Общий итог:			В интересах владельца: а)увеличить пропускную способность каналов, чтобы не терять клиентов (правда, модернизация стоит денег); б) увеличить число мест в очереди (это тоже стоит денег), чтобы задержать потенциальных клиентов.	Клиенты заинтересованы в значительном увеличении пропускной способности для уменьшения времени ожидания и уменьшения отказов.

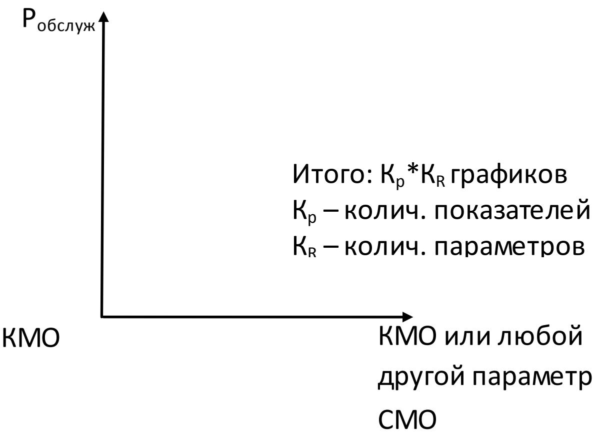
Таким образом, в таблице приведены результаты проделанного анализа существующей системы. Это дало возможность увидеть ее недостатки и определить направления улучшения ее качества. Но остаются непонятными ответы на вопросы, что именно надо сделать: 1) увеличить количество каналов или 2) увеличить их пропускную способность или 3) увеличить количество мест в очереди, и если увеличивать, то насколько? (Может быть, надо создать 3 канала с производительностью 5шт/час или 1 с производительностью 15шт/час?)

Чтобы оценить чувствительность каждого показателя качества работы СМО, надо поступить следующим образом.

Фиксируем все параметры системы, кроме одного. Затем снимаем значения всех показателей качества при нескольких значениях этого выбранного параметра (много раз моделирование для каждого значения, а потом усреднение результатов).

В итоге получаем зависимость (КМО – это количество мест в очереди в предыдущем примере):



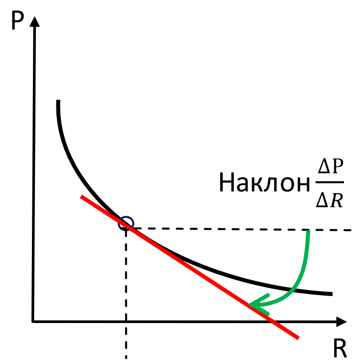


По этим кривым можно провести дополнительный анализ о перспективах движения (улучшения показателей) в ту или иную сторону. Наклон кривых хорошо показывает чувствительность, эффект от движения по определенному показателю.

т.е. если $\frac{\Delta P}{\Delta R} \to 0$, то близки к локальному оптимуму;

если $\frac{\Delta P}{\Delta R}$ большое, то резко убывает и т.д.

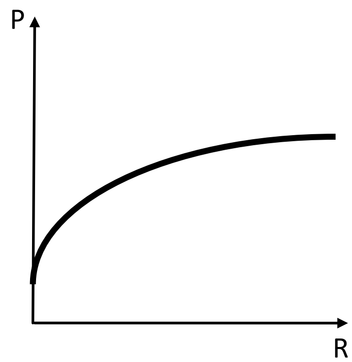
Если на основе этого анализа принимается решение об изменении какого-либо R (параметра СМО), то все графики надо переделать с учетом этого (т.е. заново все перемоделировать при этом R) для отображения новых значений P (показатели качества СМО).



Так шаг за шагом можно попытаться улучшить качество системы (так надо делать в лабораторных работах по СМО, т.е. строим графики, определяя минимум или максимум функции некоторого целевого показателя качества P, где $\frac{\Delta P}{\Delta R} \to 0$, или визуально по графику функции P от R).

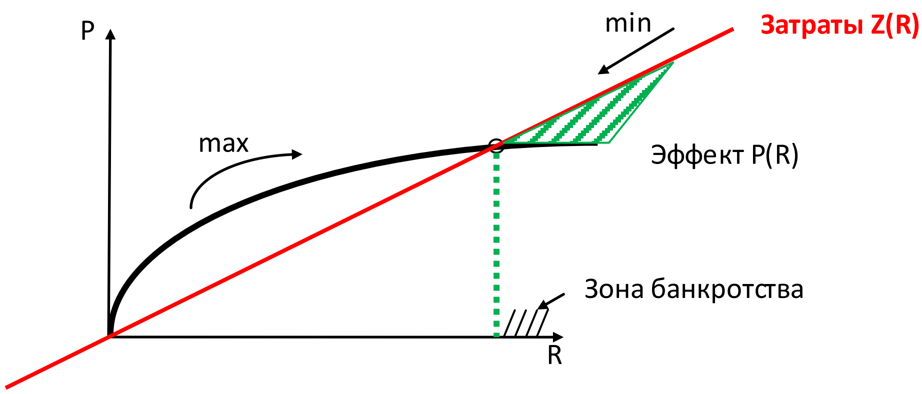
Но что делать, если кривые монотонно растут или падают? Вряд ли ответ вида $R \to \infty$ устроит владельца системы или клиента.

Параметр R – это управление; то, что находится в распоряжении владельца СМО (например, возможность заасфальтировать площадку и тем самым увеличить количество мест в очереди; поставить дополнительные колонки; увеличить поток заявок за счет увеличения затрат на рекламу и т.д.). Меняя значение параметра R, можно влиять на значение показателя качества P, цель, критерий и т.д.



Очевидно, что любое управление связано с затратами Z. И чем больше прилагаются усилия для управления, тем больше значение управляющего параметра R, и тем больше Z. Обычно полагают, что Z=CR (или exp, или $\exp^{-1}$,…).

Ясно, что когда-то сложение все новых затрат просто перестанет себя окупать. Иными словами, показатель P в сложных системах не может расти до бесконечности.



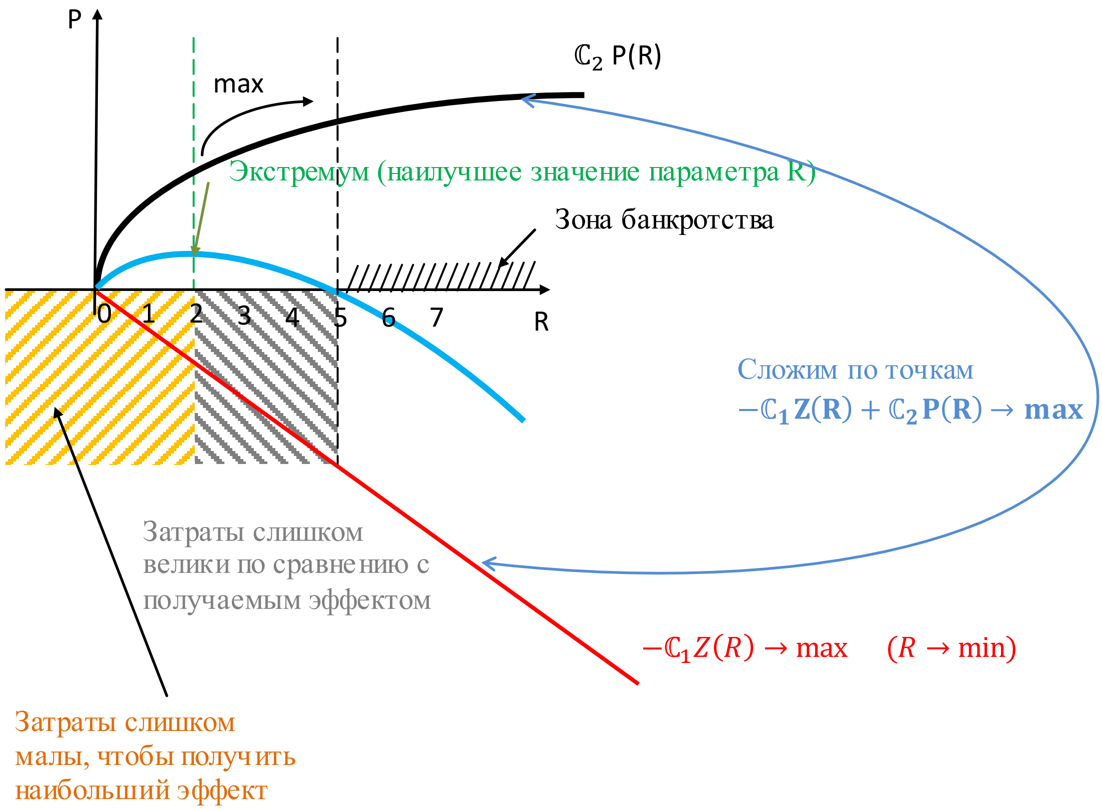
Рано или поздно его рост замедлится, а Z будет расти.

Если назначить цену за единицу затрат $R \to \mathbb{C}_1$ и цену за единицу $P \to \mathbb{C}_2$, то кривые можно:

а) сложить (если R и P надо одновременно вывести на минимум или максимум);

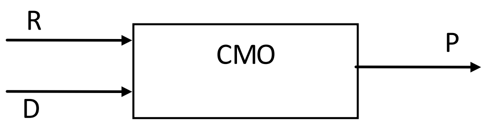
б) вычесть (из графика функции, которую выводят на максимум надо вычесть график функции, которую выводят на минимум по точкам графика)

Тогда результирующая кривая, учитывающая и эффект от управления и затраты на это, будет иметь экстремум. Значение параметра R, доставляющего экстремум функции и будет оптимальным значением $R_{опт}$.



Кроме управления R_4 показателя P в системах действуют возмущения

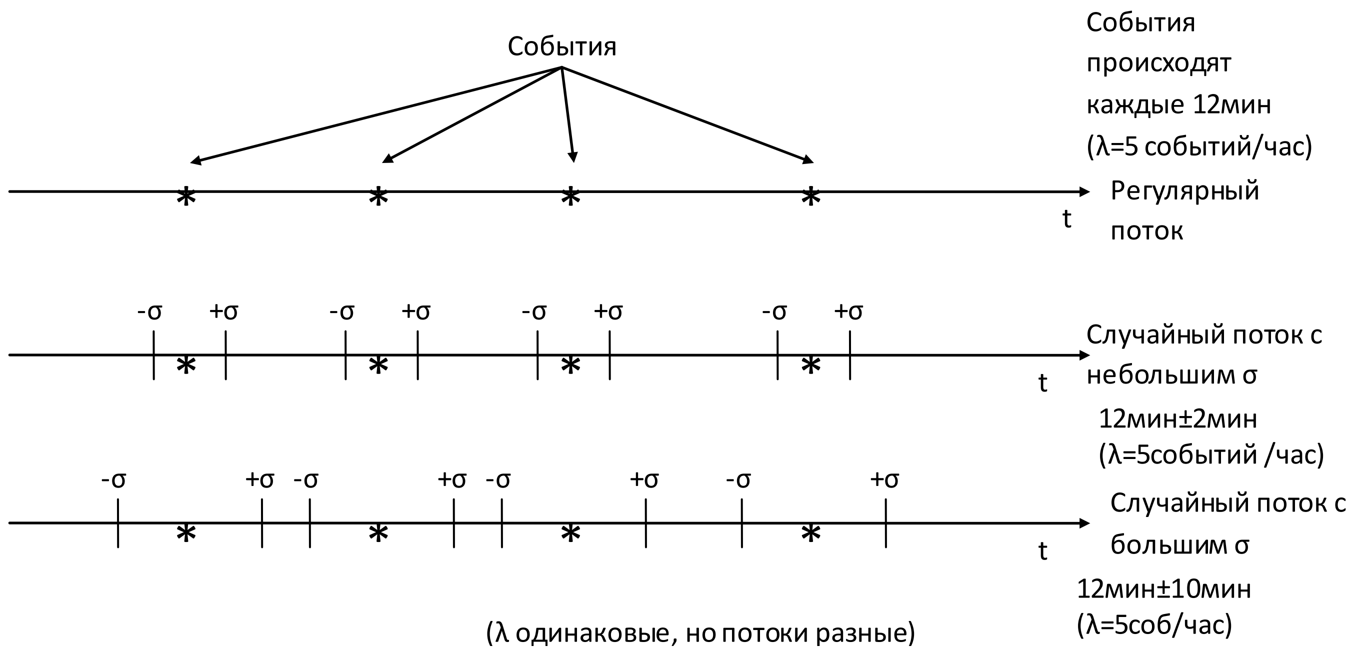
$D = \{d_1, d_2, \ldots d_N\}$. D не зависит от воли владельца системы (например, низкая



температура воздуха снижает количество клиентов). Обычно папрметр D действуют «назло» владельцу, снижая эффект показателя качества P от управляющих усилий R. ***Заметим, что реальная СМО, предоставленная сама себе, без усилий управляющего характера перестает обеспечивать цель, для достижения которой она была создана.***

V.3 Потоки Эрланга

Ранее мы рассмотрели поток Пуассона – поток без последствий: вероятность появления случайного события не зависит от момента совершения предыдущих событий.



Математическое ожидание: λ – интенсивность потока; это ожидаемое количество событий в единицу времени.

$\frac{1}{\lambda}$ – среднее время между событиями.

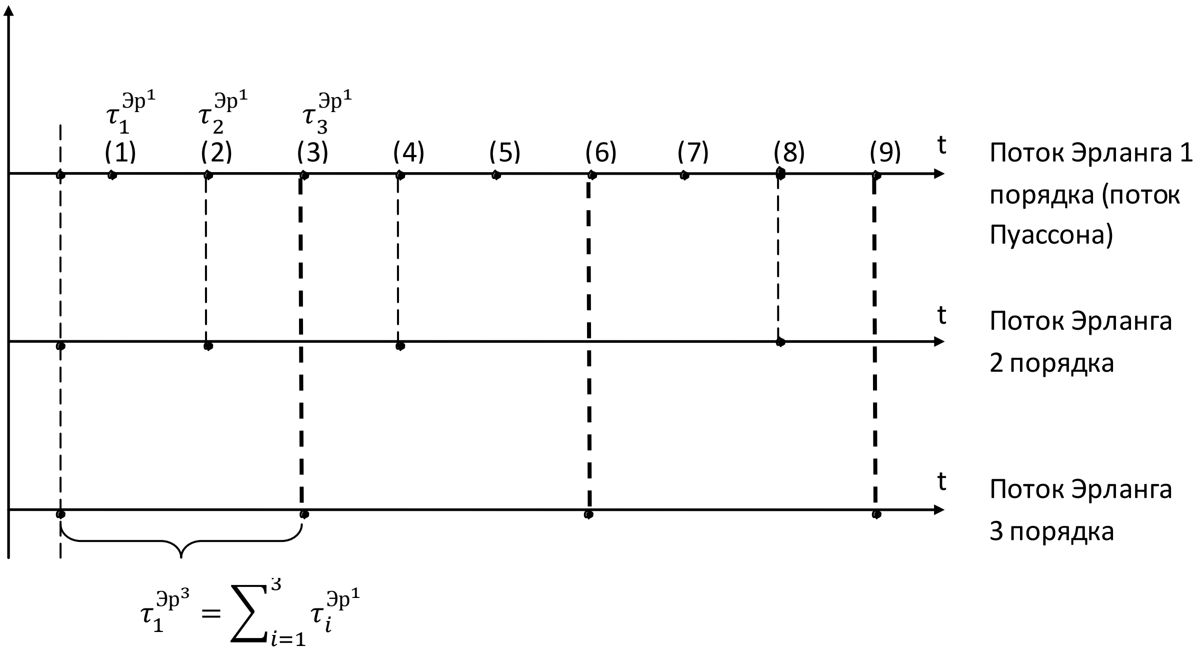
σ^2 – насколько велик разброс во время прихода событий относительно математического ожидания.

Именно дисперсия определяет случайность появления события, слабую предсказуемость моментов времени его появления.

Если предсказать момент появления каждого следующего события трудно (т.е. с точки точки зрения невозможно), то это поток без последствий; если поток детерминирован, то последствие велико – каждое событие предсказывает момент времени появления следующего события.

<u>Поток Эрланга</u> k-того порядка – это поток случайных событий, получающийся, если в простейшем случайном потоке (т.е. потоке Пуассона) сохранить каждое k-тое событие, а остальные отбросить.

<u>Порядок потока</u> – это мера последствия потока, т.е. обратной величиной к мере случайности потока является его порядок.



Просеивание событий начинает приводить к тому, что между точками появляется последствие, детерминация, которая тем выше, чем больше K (т.е. после происшествия события были какие-то незафиксированные, неважные или потерянные события). С увеличением K точки ложатся на ось времени t все более равномерно, разброс их уменьшается, регулярность растет (согласно центральной предельной теореме сумма случайных значений $\sum \xi_{\text{случ } i}$ есть величина неслучайная; чем больше сложим случайных чисел, тем предсказуемее будет их сумма ($\sum \xi_i \to N(n\mu, n\sigma^2)$, $\xi_i \succ$ одинакво(μ, σ^2))).

Значит при $k \to \infty$ события происходят в строго размеренное время, это будет регулярный поток.

$\tau^{Эр^k} = \sum_{i=1}^{k} \tau_i^{Эр^1}$ – интервал между событиями в потоке Эрланга k-того порядка.

Плотность распределения интервалов между случайными событиями в потоке Эрланга k-того порядка: $f\left(\tau_{Эр^k}\right) = \frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}$.

$\begin{cases} m_k = \frac{k}{\lambda} \\ \sigma_k^2 = \frac{k}{\lambda^2} \end{cases}$, где λ – интенсивность простейшего потока, $\frac{\lambda}{k} = \lambda_k$ - интенсивность потока Эрланга k-того порядка. $\sigma_k = \frac{1}{\sqrt{k}} \cdot \frac{1}{\lambda_k} = \frac{1}{\sqrt{k}} \cdot \frac{k}{\lambda} = \frac{\sqrt{k}}{\lambda}$ (2)

Значит, в потоках с последоствиями (которыми являются потоки Эрланга порядка выше первого) математическое ожилдание не равно дисперсии.

Таким образом, порядок потока Эрланга – это мера последствий потока.

- **Пример.**

Выход из строя лампочек уличного освещения.

$T_н$ = 100 лет. Известно, что среднее время работы изделия на отказ составляет 1,5 года; среднеквадратичное отклонение составляет 0,5 года.

Т.е. заданы $M_к = 1,5$; $\sigma_к = 0,5$.

Т.к. $M_к \neq \sigma_к$ => $k \neq 1$ – это поток с последствиями. Интенсивность этого потока: $\lambda_k = \frac{1}{M_k} = 0{,}67$ - т.е. в течение года в среднем перегорает 0,67 лампочки или 67 штук за 100 лет. Из (1) => $k = \frac{1}{\sigma_k^2} / \lambda_k^2 \approx 9$. (Т.е. при моделировании будем брать каждое девятое событие из простейшего потока Пуассона, это обеспечивает достаточно высоую детерминированность потока). Вычислим интенсивность порождающего потока Пуассона: $\lambda = \lambda_k \cdot k = 6$.