Конспект лекций по курсу «ИССЛЕДОВАНИЕ ОПЕРАЦИЙ»

Бакалавры 2020/21 уч.год.

Бакалавры. **3-й курс**, весна, **6-й семестр**, 3 зач.ед., 108(108)час: аудит (50: **лекц-34**, ПЗ-0, **ЛР-16**), СРС – 58. **Экзамен**.

ЛИТЕРАТУРА

1. Гермейер Ю.Б. Введение в теорию исследования операций. М.: Наука, 1971, 384с.
2. Вагнер Г. Исследование операций. В трех томах. М.: Мир, 1973.
3. Вентцель Е.С. Исследование операций. М.: Советское радио, 1972 г.
4. Вентцель. Е.С. Исследование операций: задачи, принципы, методология. – М.: Наука 1988 – 208с.
5. Дегтярев Ю.И. Исследование операций. Учебник. М.: Высшая школа, 1986 г, 320 с.:ил.
6. Дегтярев Ю.И. Системный анализ и исследование операций М.: Высшая школа, 1996 г, 335 с.:ил.
7. Таха Х. Введение в исследование операций в 2-х книгах. М. Мир, 1985, 496с.
8. Бомас В.В., Павленко А.И., Росин М.Ф. Оценка эффективности решений в АСУ ЛА по многим критериям. М.: МАИ, 1989.-46с.
9. Росин М.Ф. Принципы и модели принятия решений в АСУ ЛА. М.: МАИ, 1994 г.
10. Бомас В.В., Булыгин В.С. Элементы теории Марковских процессов и ее технические приложения. Учебное пособие. М.: МАИ, 1980. 84с.
11. Бомас В.В., Красовская М.А., Ескин В.И. Лабораторные работы по курсу «Исследование операций». М.: МАИ, 1992. – 28 с.:ил.

12. Кини Р.Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения: М.: Радио и связь, 1981 г.
13. Ларичев О.И. Объективные модели и субъективные решения. – М.: Наука, 1987
14. Саати Т., Принятие решений. Метод анализа иерархий, М.: «Радио и связь», 1993 г.
15. Трахтенгерц Э.А. Компьютерная поддержка принятия решений. М.:СИНТЕГ, 1998
16. Бомас В.В., Судаков В.,А., Афонин К.А., Поддержка принятия многокритериальных решений по предпочтениям пользователя. СППР DSS/UTES, М.: Изд-во МАИ, 2008 г.
17. Павленко А.И. Формализация задач принятия решений и выбора. Учебное пособие. М.: МАИ-ПРИНТ, 2009. -88с.:ил.
18. Хемди А. Таха. Введение в исследование операций = Operations Research: An Introduction. — М.: Вильямс, 2007. — 912 с. — ISBN 0-13-032374-8
19. Грешилов А. А. Математические методы принятия решений. — М.: МГТУ им. Н.Э. Баумана, 2006. — 584 с. — ISBN 5-7038-2893-7
20. Дунин – Барковский И.В., Смирнов Н.В. теория вероятностей и математическая статистика. – М.: Гостехтеориздат, 1955.
21. Хахулин Г.Ф. «Основы конструирования имитационных моделей» Учебное пособие. М.: НПК «Поток», 2002.
22. Полляк Ю. Г. “Вероятностное моделирование на ЭВМ” 1971.
23. Шеннон Р. “Имитационное моделирование искусство и наука” Мир 1978.
24. Клейнен Д. “Статистические методы в имитационном моделировании” 1978.
25. Бусленко Н. П. “Моделирование сложных систем” Москва Наука 1978.

СОДЕРЖАНИЕ

Лекция 1

ЦЕЛИ И ЗАДАЧИ «ИССЛЕДОВАНИЯ ОПЕРАЦИЙ»

В общем случае под операцией понимается любая целенаправленная деятельность. Т.к. способов ее осуществления, как правило, достаточно много, то результатом исследования должен быть выбор одного из них. Поэтому исследование операций заключается в сопоставлении и упорядочении различных способов (вариантов) их проведения (*ранжирование*), или в поиске наилучшего варианта (*оптимизация*). Заключительным этапом исследований является принятие решения о том, какой из возможных способов проведения операции следует применить.

"**Исследование операций**" (**ИО**) как самостоятельная научная дисциплина насчитывает более 70 лет. Впервые это название появилось в годы второй мировой войны, когда в вооруженных силах некоторых стран (США, Англия) были сформированы специальные группы научных работников (физиков, математиков, инженеров), в задачу которых входила подготовка проектов решений для командующего боевыми действиями. Эти решения касались главным образом боевого применения оружия и распределения сил и средств по различным объектам. Подобного рода задачами занимались и в нашей стране. В дальнейшем **ИО** распространило область применения на различные области практики: промышленность, сельское хозяйство, транспорт, связь, торговля, бытовое обслуживание и т. д.

Ключевыми понятиями «Исследования операций» являются: критерий эффективности, математическая модель, оптимальное планирование, компьютерное моделирование, принятие решений.

Ряд этих понятий как, например, оптимальное планирование и компьютерное моделирование являются содержанием самостоятельных научных дисциплин, которые изучаются в отдельных курсах.

Кроме того, в задачах, требующих обработки больших информационных массивов, поиск оптимальных или рациональных решений практически невозможен без использования современных информационных технологий и базирующихся на них *систем поддержки принятия решений* (СППР). Поэтому методы исследования операций должны ориентироваться на их использование.

Рассмотрим основные понятия и определения.

Под исследованием операций будем понимать "применение математических, количественных методов для обоснования решений во всех областях целенаправленной человеческой деятельности".

Операцией называется совокупность взаимосогласованных действий, направленных на достижение вполне определенных целей. Операция всегда есть управляемое мероприятие, т. е. От нас зависит каким образом выбрать те или иные параметры, характеризующие ее организацию, протекание. Например: набор технических средств, применяемых в операции, порядок их применения и т.д.

Если ***цель операции*** определена и существуют различные пути ее достижения, то желательно найти лучший из них. При этом понятие "лучший" имеет смысл только тогда, когда может быть назван показатель (или группа показателей) по которым можно сравнить различные варианты. Например, монтаж оборудования в цехе с целью обеспечения выпуска продукции в заданный срок; обеспечение производства сырьем при минимальных расходах на перевозки; распределение средств разведки между районами поиска, с тем, чтобы обеспечить максимум вероятности обнаружения подводных лодок противника и т. п.

Лучшим (оптимальным) считается такой способ действия, который в наибольшей степени способствует достижению поставленной цели с точки зрения выбранного показателя – критерия эффективности.

Критерий эффективности операции – показатель (или группа показателей), количественно отражающий цель операции, позволяющий сопоставить между собой результаты предпринимаемых действий и цель операции. Используется как для сравнения между собой различных вариантов действий во время операции, так и для характеристики результатов, полученных по окончании операции.

Оперирующая сторона (ОС) – отдельные лица или группа лиц, активно стремящиеся в рамках данной операции к достижению поставленной цели.

Противодействующая сторона (ПС) – другая сторона, участвующая в операции и преследующая в ней свои (как правило, противоположные относительно ОС) цели. Например, противник в военных действиях. Если такой стороны, принимающей активные действия в операции, нет, то её, как правило, не учитывают.

Активные средства (АС) проведения операции – совокупность материальных, энергетических, финансовых, трудовых и др. ресурсов, а также организационных возможностей, используемых оперирующей стороной для обеспечения успешного хода операции. ОС должна обладать определенной свободой выбора активных средств и возможностью разнообразного их использования. Например: АС – производственные мощности, фонд з\платы, трудовые ресурсы, совокупность боевых средств, вычислительные ресурсы ЭВМ и пр. АС характеризуются своим количеством и параметрами, определяющими эффективность их применения.

Стратегии (для ОС) – допустимые способы расходования (использования) оперирующей стороной имеющихся в ее распоряжении активных средств (допустимые – в смысле физической реализуемости). То есть *правила поведения*

ОС в операции по расходованию ею АС. Например, размещение оборудования в цехе; распределение ФЗП трудовых ресурсов во времени между различными видами работ, распределение поисковых единиц между районами поиска п/л и т. д.

Среди всех возможных стратегий, очевидно, находятся и оптимальные. Кроме стратегий на результат операции влияют действующие факторы, определяющие условия, в которых протекает операция.

Действующие факторы – совокупность объективных условий и обстоятельств, определяющих особенности операции, действующие во время ее протекания и определяющие ее исход. Совокупность действующих факторов характеризуют обстановку, в которой проводится операция. Различают контролируемые и неконтролируемые факторы.

Контролируемые факторы находятся в распоряжении оперирующей стороны.

Неконтролируемые факторы, в зависимости от информированности о них исследователя операции, следует разделить на:

- ***Определенные*** - значения которых заранее, (до операции) известны;
- ***Неопределенные*** - значения которых заранее не известны. Среди *неопределенных факторов* необходимо выделить:
  - *статистически определенные факторы* – значения которых во время подготовки операции исследователям не известны, но они обладают свойствами повторяемости и статистической устойчивости, т.е. случайные величины (процессы) с известными законами распределения;
  - *статистически неопределенные факторы* – для которых известна только область их возможных значений, но не известен точно закон их распределения. Неопределенность подобного рода может возникать либо за счет наличия субъектов (объектов), действующих независимо от ОС (или даже ей вопреки). Такие факторы можно назвать ***стратегиями противника***

(противодействующей стороны), либо из-за недостаточной изученности каких-либо процессов – их называют ***природными факторами*** (например, известны математическое ожидание и дисперсия, но неизвестен вид закона распределения), либо за счет нечеткого знания цели операции или критерия эффективности (например, неопределенность выбора критерия эффективности предприятия, выпускающего продукцию существенно различных типов).

Для того, чтобы для исследования операции можно было применять формальный математический аппарат, необходимо построить математическую модель операции.

Математической моделью операции называются всякие формальные соотношения, устанавливающие количественную связь принятого критерия эффективности, отражающего цель операции, с действующими факторами. Очевидно, что чем удачнее будет подобрана математическая модель, чем лучше она будет отражать характерные черты операции, тем успешнее будет исследование и полезнее вытекающие из него рекомендации. *Математические модели могут иметь вид* формул, систем уравнений, неравенств, таблиц, алгоритмов и т.п., отражающих количественную связь между критерием эффективности операции и параметрами, отражающими действующие в операции факторы.

На основе разработанной математической модели операции *производится* ***поиск решения*** – выбор определенных значений управляемых оперирующей стороной факторов (стратегий), среди возможных, зависящих от ее выбора. Решения, наилучшим образом отвечающие целям операции и предпочтительнее с точки зрения выбранного критерия эффективности называются ***оптимальными решениями***.

При поиске решений необходимо различать два аспекта:

- подготовка решений (или рекомендаций по выбору решений);
- принятие (выбор) окончательного решения.

<u>Подготовкой решений</u> занимается *<u>исследователь операций</u>* – специалист или группа специалистов (коллектив), осуществляющий**:** - разработку стратегий, допустимых в операции; - формализацию задачи; - выбор критерия эффективности; - разработку математической модели (или моделей) операции и - отыскание оптимальных стратегий среди множества возможных. Исследователь входит в состав оперирующей стороны.

<u>Принятие решения</u> состоит в выборе решения, которое будет реализовано в ходе операции. Оно выходит за рамки исследования операций и относится к компетенции ответственного **<u>лица</u>** или группы лиц, **<u>принимающего решение (ЛПР)</u>**, которому предоставляется право окончательного выбора и на которого возлагается ответственность за этот выбор.

Делая такой выбор, наряду с рекомендациями, вытекающими из математических расчетов, проведенных на разработанных математических моделях, ЛПР может учитывать ряд дополнительных соображений количественного и качественного характера, неучтенных в расчетах в виду несовершенства математического аппарата и из-за сложности их формализации (например, характер противника и т.п.).

Непременное участие лица, принимающего решения даже в полностью автоматизированной системе управления, принимающей решение без непосредственного участия человека, обеспечивается фактическим выбором одного из вариантов управляющего алгоритма, по которому осуществляется работа системы. Кроме того в автоматизированных системах управления подобного рода обычно предусматривается возможность вмешательства человека в ход управляемого процесса.

Таким образом, в качестве ***<u>основных этапов исследования операций</u>*** можно сформулировать следующие:

- <u>Выбор критерия эффективности</u> (одного или нескольких);

- Определение совокупности активных средств, управляемых ОС и формулировка стратегий – допустимых способов их использования;
- Выявление основных факторов, действующих во время протекания операции;
- Построение математической модели операции;
- В рамках принятой модели операции нахождение оптимальных (наилучших) решений, отвечающих экстремальным значениям выбранного критерия эффективности среди всех возможных решений.

Критерий эффективности операции.

Как уже говорилось, ***критерий эффективности операции*** – это показатель (или группа показателей), количественно отражающий цель операции, позволяющий сопоставлять между собой результаты предпринимаемых действий и цель операции.

Они должны отвечать следующим основным требованиям:

- представительность;
- критичность к исследуемым параметрам;
- максимально возможная простота;
- объединение в себе, по возможности, всех основных элементов исследуемой операции;
- правильный учет стохастических факторов.

Рассмотрим эти требования подробнее.

Представительность. Критерий должен позволять оценивать эффективность основной задачи операции, а не второстепенной. Неправильный выбор критерия приводит к тому, что все исследования, проведенные на его основе, оказываются бесполезными. А полученные решения – нецелесообразными, а подчас вредными и приводящими к неоправданным затратам и потерям. *Например*: пресловутый “Вал” (объём выпущенной продукции) в качестве основного критерия оценки хозяйственной деятельности предприятий; минимальная стоимость работ при участии в конкурсе проектов; максимальная прибыль – не всегда соответствует высокому качеству разработки.

Критерий должен быть ***критичным*** к исследуемым параметрам, т.е. чувствительным к их изменениям. Чем эта критичность выше, тем лучше.

Критерий должен быть ***простым***, т.к. введение в него второстепенных составляющих может усложнить исследования, не приводя к уточнениям получаемых выводов, не позволяя выявить основных результатов (эффектов) операции, как говориться “за деревьями не видно леса”.

Желательно, чтобы критерий был ***единым***, т.к. решение задачи исследования и поиска “наилучших” решений при наличии двух и более критериев крайне затруднено и без введения дополнительных предположений, условий не дает возможности использования строгих математических методов. Однако следует отметить, что сведение задачи к одному критерию не всегда позволяет в полной мере учесть все цели, стоящие перед операцией, что значительно обедняет задачу и не всегда возможно. В этом случае приходим к необходимости формирования ***векторного критерия эффективности*.** Проблемы поиска решений в этом случае представляют собой особую задачу и будут рассматриваться отдельно.

При учете стохастического характера исследуемых процессов при выборе критерия эффективности необходимо учитывать возможность повторяемости операции, допустимость осреднения ее результатов по многим реализациям, характер неопределенности случайных факторов и т. п.

Лекция 2.

*Ограничимся, для начала, рассмотрением задач исследования операций с одним (**скалярным**) критерием эффективности.*

Обозначим:

- **W** - **критерий эффективности** (показатель эффективности) операции;
- **a** - совокупность действующих в операции факторов, неуправляемых ни одной из сторон, значения которых известны заранее, детерминированы и в ходе операции не изменяются (заданные **параметры** АС, внешней среды);
- **x∈X** – совокупность факторов, управляемых оперирующей стороной, характеризующих ее поведение во время операции и составляющих в совокупности решение (X – множество всех возможных решений ОС - **стратегии оперирующей стороны**);
- **y∈Y** – совокупность факторов, управляемых противодействующей стороной, где Y – множество всех возможных решений ПС - **стратегии противодействующей стороны** (если необходимо учитывать возможность её различного поведения в операции);
- **z∈Z** – множество факторов, не контролируемых ни оперирующей, ни противодействующей стороной, значения которых заранее не известны (***неконтролируемые факторы***).

Тогда значение, которое примет показатель эффективности **W** в результате операции зависит от всех этих групп факторов и может быть записано как некий функционал

$$\mathbf{W = W(a, x, y, z)}$$

В зависимости от цели, стоящей перед операцией можно различить два вида критериев эффективности: “*качественные*” и “*количественные*”.

В случае **качественного критерия эффективности** целью операции является достижение определенного заданного результата (эффекта), который может быть получен или не получен, и, соответственно, поставленная цель является выполненной, если заданный эффект достигнут, и не выполненной – в противном случае. Например, изготовление работоспособного изделия, успешная (безаварийная) посадка ЛА, поражение объекта противника и т.д.

В этом случае результат операции может быть оценен двумя числами:

1 – если цель операции достигнута;

0 – если цель операции не достигнута.

Если все факторы, влияющие на операцию детерминированные, то

$$\mathbf{W = W(a, x^*, y^*,) =} \begin{cases} \mathbf{1} \text{ – если желаемый результат достигнут;} \\ \mathbf{0} \text{ – если результат операции не достигнут.} \end{cases}$$

Здесь **x*, y*** - фиксированные значения управляемых факторов.

Если же на ход операции оказывают влияние случайные факторы Z, то значение показателя W = W(a, x*, y*, z*) также будет случайно и использование его в качестве критерия эффективности в этом случае невозможно.

В случае **количественного критерия эффективности** целью операции является получение наилучшего значения некоторой величины, оценивающей конечный эффект операции. Например, объем выпускаемой продукции, время безотказной работы оборудования, число пораженных объектов противника и т.д.

Если при этом все факторы, влияющие на результат операции, детерминированные, то в качестве критерия эффективности можно использовать значение этой величины.

$$\mathbf{W = W(a, x^*, y^*).}$$

При наличии случайных, статистически определенных факторов **Z**, влияющих на результат операции, в качестве критерия эффективности можно принять, например, *математическое ожидание* некоторой случайной величины **R**, характеризующей результат операции

$$\mathbf{W = M_z\,[R(a, x^*, y^*, z)] = W(a, x^*, y^*),}$$

или каких-либо других величин, полученных с учетом случайности R.

При известном законе распределения f(z) случайных факторов Z значение W может быть найдено как

$$\mathbf{W(a, x^*, y^*) = \int\limits_{z \in Z} R(a, x^*, y^*, z)\, f(z)\, dz\ ,}$$

где **f(z)** – плотность распределения вероятностей случайных факторов **Z**.

Как видим, неопределенность значений случайной величины **R**, которые она примет в ходе операции при случайных значениях величины **Z**, при этом устраняется.

Следует заметить, что применение в качестве критерия эффективности математического ожидания *возможно* только тогда, когда допустимо осреднение результатов операции либо при многократном ее повторении. *Например*: - среднее время безотказной работы оборудования при многократной его эксплуатации), либо, если осреднение результатов может быть осуществлено в ходе одной операции за счет большого числа одновременно выполняемых однотипных действий, (число “хороших” изделий, выпускаемых на 100 станках автоматах в течение смены). Так как “недостача” показателя в одном случае может быть компенсирована “избытком” в другом; - суммарный доход в случае многократного повторения операции.

Если осреднение результатов по многим операциям недопустимо или имеется большой разброс величины **R** относительно среднего, то использование в качестве критерия эффективности математического ожидания может привести при приемлемых средних показателях к недопустимым результатам в одной отдельной операции (например, среднее число успешных полетов в авиации).

В этих случаях в качестве показателя эффективности **W** можно использовать вероятность того, что случайная величина R, характеризующая результат операции, достигнет значения, не менее заданного:

$$\mathbf{W(a, x^*, y^*) = P_z\,[R(a, x^*, y^*, z) \geq R_{зад}],}$$

или иначе

$$\mathbf{W(a, x^*, y^*) = \int\limits_{(z \in Z:\ R(a,\, x^*,\, y^*,z) \geq R_{зад}} f(z)\, dz}$$

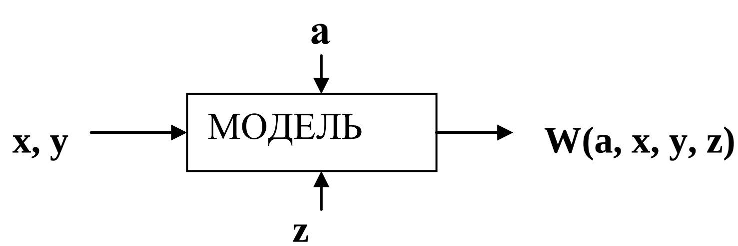
В качестве случайной величины **R**, в частности, можно использовать и качественный показатель эффективности, характеризующий факт достижения (1), либо нет (0) заданной цели операции: поражение (не поражение) объекта противника.

$$\mathbf{W(a, x^*, y^*) = \{1,0\}).}$$

В этом случае возможно использование математического ожидания данного показателя (если осреднение по операциям возможно), либо вероятности достижения эффекта в отдельной операции. Это позволяет учесть неопределенность исхода операции при наличии случайных факторов, перейдя, таким образом, от качественного показателя эффективности к количественному.

Математические модели операций.

Как уже говорилось выше, под математическими моделями операции будем понимать всякие формальные соотношения, устанавливающие количественную (качественную) связь между: - управляемыми переменными **x∈X** (**y∈Y**); - неизменяемыми в ходе операции факторами, техническими параметрами элементов систем и устройств, используемых в операции **a**; - неконтролируемыми факторами **Z**, и показателем (показателями) эффективности операции **W**, характеризующим ее результат. Т.е.



Для построения математической модели операция упрощается, схематизируется и описывается с помощью того или иного математического аппарата.

В самых простых случаях операция описывается простыми алгебраическими уравнениями. В более сложных случаях, когда требуется рассматривать явление в динамике, применяется аппарат дифференциальных уравнений. Такие модели, как правило, являются ***детерминированными***. В них ход операции определяется как в настоящем, так и в будущем. Такие модели широко используются при планировании перевозок, в теории расписаний, распределении ресурсов и т.д.

В случаях, когда при описании хода операции необходимо учитывать влияние случайных факторов, оценивать результаты операции с учетом вероятностей тех или иных событий используют ***вероятностные модели***.

К ним относятся, например, модели массового обслуживания, описывающие операцию как потоки обслуживания случайных требований в исследуемой системе. Кроме того вероятностные модели широко применяются при описании военных операций и т.п.

В наиболее сложных случаях, когда развитие операции и ее исход зависят от большого числа, сложно переплетающихся между собой, факторов, часть из которых случайные (неопределённые), и аналитические методы исследования вообще отказываются служить, широкое применение находит метод ***имитационного моделирования***, суть которого заключается в следующем.

Процесс развития операции с учетом всех случайных факторов, подлежащих учету, воспроизводится (моделируется) на ЭВМ. В результате получается одна “реализация” случайного протекания операции со случайным ее исходом. При многократном повторении хода операции, за счет ее случайности, получим множество различных реализаций, обработав которые как обычный статистический материал можно найти средние характеристики процесса, получить представление как влияют на результат операции те или другие параметры и управляемые переменные. Такие модели часто называют имитационными.

Особо необходимо выделить так называемые **игровые модели**. Они позволяют описывать и изучать операции, в которых каждая из сторон, участвующих в операции (**игре**), придерживается своих целей, стараясь, по возможности, получить информацию о поведении противника, которое заранее не известно, извлечь выгоду из его ошибок и т.п. Примером такой операции является игра двух сторон (игроков), в которой интересы каждого из игроков прямо противоположны, набор возможных решений сторон известен, информация о поведении противоположной стороны в игре отсутствует (так называемая **игра с ну-**

левой суммой). Изучением таких операций занимается специальный раздел ИО – теория игр.

При разработке математическая модель должна отвечать ряду общих требований.

1) Модель должна быть **адекватной**, т.е. несмотря на некоторые неточности отображения исследуемой операции (системы) способна обеспечить достаточно хорошее совпадение результатов, полученных с помощью модели с результатами, которые были бы получены при тех же исходных данных на реальной системе, что позволяет быть достаточно уверенным в правильности определения результатов операции в других условиях. Это дает возможность использования рекомендаций, выработанных на основе применения математических моделей. Вопрос оценки адекватности модели достаточно сложен, особенно в случаях, когда нельзя сравнить результаты счета модели с работой реальной системы (например, в виду отсутствия таковой).

2) Математическая модель с одной стороны должна **учитывать** все существенные **факторы**, **влияющие на ход операции**, но с другой стороны, быть по возможности простой, не “засоренной” массой мелких деталей (второстепенных), которые лишь усложняют исследования и затрудняют выявление основных закономерностей, делая подчас их применение для анализа операций практически невозможным.

3) Модель должна быть **чувствительной** к изменениям основных параметров операции (системы), что позволяет использовать ее для исследования их влияния на найденные решения.

4) Кроме того точность (**детальность**) модели должна быть соизмерима с точностью информации о значениях параметров **a** и факторов **Z**, которой мы располагаем. В противном случае мы можем получать недостоверные результаты даже при, казалось бы, чрезвычайно подробной, детальной модели.

Общих рекомендаций для построения математических моделей не существует.

Лекция 3.

Разновидность задач «Исследования операций»

Все задачи «Исследования операций» можно условно разбить на две категории: *прямые и обратные*.

Прямые задачи отвечают на вопрос “что будет, если в заданных условиях будет принято какое-то конкретное решение **x*∈X**. В частности – при заданном решении **x*** и заданных детерминированных значениях неуправляемых факторов **a**, чему будет равен выбранный показатель эффективности операции **W(a, x*)** или ряд показателей $\overline{W} = \overline{W}$ **(a, x*)** ***при векторном критерии эффективности*** **.**

Значения неизменяемых в ходе операции факторов **(y, z)** *в дальнейшем будем опускать, чтобы не загромождать запись.*

Для ответа на этот вопрос необходимо построить математическую модель операции, позволяющую выразить один или несколько показателей эффективности **W** через заданные условия **a** и элементы решения **x**.

Решение прямых задач, как правило, достаточно простое. Однако, в случаях при анализе операций с большим числом взаимосвязанных факторов, (например, в задачах массового обслуживания) их решение весьма сложно и требует разработки специальных методов и приемов.

Обратные задачи отвечают на вопрос – какое необходимо выбрать решение **x∈X** (оптимальное или рациональное) для того, чтобы обеспечить *наилучшее* значение показателя эффективности **W**.

В случае ***скалярного критерия эффективности W,*** под *наилучшим* будем понимать максимальное значение показателя. Тогда обратная задача ИО может быть записана следующим образом:

При заданном комплексе условий **a,** найти такое решение **x = x°** , **(x∈X),** которое обращает показатель эффективности операции **W** в максимум

$$\mathbf{W\,(x^\circ) = \max_{x \in X} \; W\{a, x\}\;.}$$

В случае минимизации показателя **W**, обуславливаемом его физическим смыслом (например, расходы на создание системы, минимизации веса ЛА), задача формально сводится к задаче максимизации путем изменения знака показателя **W** на противоположный.

Если число возможных вариантов решения, образующих множество **X** невелико, то для поиска оптимального решения необходимо, для каждого из возможных значений **x∈X**, , решить прямую задачу и определить соответствующее значение показателя **W**. Сравнив их между собой, можно указать одно или несколько решений, при которых **W** достигает максимума. Такой способ называется ***простым перебором***.

В случае, если множество возможных вариантов решения **X** велико, то поиск среди них оптимального простым перебором бывает весьма затруднителен, а подчас, просто невозможен. В этих случаях применяют, так называемый, метод направленного перебора, заключающийся в том, что оптимальное, или близкое к нему решение находится в результате ряда, специальным образом организованных шагов, в результате которых каждое из последующих решений приближает нас к искомому оптимальному.

В этих случае, задача поиска оптимального решения относятся к классу задач ***математического программирования***. (Термин «программирование» от английского programming – составление плана или программы действий, здесь следует понимать в смысле “поиска наилучших планов, решений”, а не в смысле разработки программного обеспечения для ЭВМ.)

В ***<u>случае векторного критерия эффективности</u>*** $\overline{W}$, при решении обратной задачи, ***<u>под наилучшим (рациональным)</u>*** обычно понимается решение, обеспечивающее:

1) <u>Максимальное (минимальное) значение некоторого обобщенного</u> (скалярного) <u>показателя **U**</u>, который представляет собой результат формализованной (неформализованной) свертки векторного критерия $\overline{W}$. То есть

$$\max U = U(\overline{W}) .$$

В этом случае (свертка векторного критерия) основная проблема заключается в построении обобщенного показателя **U**, отвечающего цели операции и учитывающего различный вклад частных показателей вектора $\overline{W}$.

<u>Например</u>,

$$\mathbf{U = \sum_i \alpha_i W_i \quad \nu \le \alpha i \le 1 , \quad \sum_i \alpha_i = 1 \;\; (i = 1, k) .}$$

Здесь весовые коэффициенты α_i характеризуют вклад частных показателей **Wi** в общий. На их значения, как правило, накладываются ограничения:

$$\mathbf{0 \le \alpha_i \le 1 , \qquad \sum_i \alpha_i = 1 \qquad (i = 1, k),}$$

где **k –** число компонент векторного критерия эффективности $\overline{W}$.

Поиск решения при выбранной функции свертки U($\overline{W}$) аналогичен поиску оптимального решения по скалярному критерию эффективности.

2) <u>Решения (одно из решений), отвечающее условию</u>, что нельзя найти другое решение, позволяющее улучшить любой из показателей **Wi (i = 1,** k**)** не ухудшая при этом значения других показателей **W**ξ **(**$\xi \ne$**i)** (хотя бы одного из них). Такие решения называются ***<u>эффективными</u>*** <u>или</u> ***<u>Паретовскими</u>*** ***(***$x_n \in Xn$***).***

*П*оиск эффективных (Паретовских) решений $\mathbf{x_п \in X_п}$ может быть осуществлен путем прямого перебора всех возможных решений $x \in X$ (если оно конечно) одновременно с процедурой сравнения отдельных компонент векторного критерия **W**$_i$ **(i = 1,** k**).**

Если множество X бесконечно или велико, и прямой перебор невозможен, то целесообразно выработать некоторое правило, по которому можно осуществлять целенаправленный поиск только, по крайней мере, части точек $\mathbf{x \in X_п}$ (например, переходя от непрерывного множества значений **X** к дискретному с некоторым шагом **Δx**), исключая при этом из рассмотрения заведомо неперспективные точки.

Следует подчеркнуть, что выбор наилучшего решения среди множества найденных Паретовских $\mathbf{x \in X_п}$ производится ЛПР.

Таким образом, задача поиска рационального решения даже при векторном критерии эффективности в ряде случаев также сводится к необходимости решения ***задачи*** (или совокупности задач) ***математического программирования***.

СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ (СМО). ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

При исследовании операций часто приходится сталкиваться с работой систем, в которых в той или иной форме производится «обслуживание» **потока** однородных **требований** (заявок).

Например: АТС, магазины, различные технические системы, системы военного назначения и т.п.

Такие системы обычно рассматривают как **Системы Массового Обслуживания (СМО).**

Каждая СМО состоит из какого-то числа обслуживающих единиц, которые называются **каналами обслуживания**. *Например:* линии связи, продавцы, пункты технического обслуживания и т.п.

По числу каналов обслуживания СМО могут быть одноканальными и многоканальными.

Требования на обслуживание или заявки поступают в систему в какие-то случайные моменты времени. **Обслуживание** (выполнение) заявки продолжается, в общем случае, также какое-то случайное время $T_{обс.}$, после чего канал освобождается и готов к приему следующей заявки. Случайный характер потока заявок и времени обслуживания приводит к тому, что в какие-то периоды времени на входе СМО скапливается излишне большое число заявок, и они либо становятся в **очередь**, либо покидают систему **необслуженными**. В другие периоды времени система может работать с недогрузкой, а то и вовсе простаивать.

Например:

1. **АТС** – многоканальная система, в которую поступают вызовы от абонентов на обслуживание (соединение с другим абонентом). Если каналы (ли-

нии связи) занят, то при наборе первых [трех] цифр возникает сигнал «занято» и заявка получает отказ. Очередь на обслуживание отсутствует.

2. **Обслуживание М станков-автоматов N рабочими (N≤ M).** В их обязанности входит контроль, наладка станков, заправка их сырьем, устранение неисправностей и т.п. Каждый рабочий одновременно может обслуживать только один станок. Если число отказавших станков больше числа рабочих (каналов обслуживания), то образуется очередь на обслуживание.
3. **Система передачи сообщений** (*например*: РТС, селектор). В этом случае возможно ожидание на входе в систему – образование очереди, максимальная длина которой определяется техническими возможностями системы и ограничена. Кроме того, в зависимости от важности сообщений, они могут обслуживаться как в порядке поступления, так и вне очереди (с приоритетом) прерывая при ней идущие сообщения.

Кроме того в реальных системах могут встречаться и другие условия функционирования систем. *Например:* ограничение времени нахождения заявки в очереди, поступление заявок группами и т.д.

Предметом теории массового обслуживания (ТМО) является построение и исследование математических моделей, связывающих заданные условия работы СМО (число каналов обслуживания, их производительность, правила обслуживания и т.п.) с интересующими нам показателями эффективности ее работы (среднее число заявок, обслуживаемых системой за указанный период, среднее число занятых каналов, средняя длина очереди, вероятность обслуживания заявок в системе, среднее время ожидания обслуживания и т.п.). Т.е. аппарат ТМО позволяет решать прямые задачи исследования операций.

Процесс функционирования СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем. Состояние системы (число занятых каналов, число заявок в очереди и т.п.) изменяется скачкообразно в моменты появления каких-либо событий – приход очередной заявки, окончание обслуживания заявки и т.п.

Качество функционирования СМО определяется успешностью обслуживания требований входящего потока. Наиболее важные и наиболее общие показатели – это полнота и скорость обслуживания. Именно они характеризуют пропускную способность СМО. Способы вычисления этих показателей зависят от характера входящего потока (внешние характеристики) и от типа СМО (внутренние характеристики), и определяют модель конкретной СМО.

Математический анализ СМО существенно упрощает, если процесс ее работы Марковский. **Процесс** называется **Марковским**, если вероятностные характеристики его в будущем зависят только от состояния процесса в настоящем. При этом не важно, каким путем оно в это состояние пришло, - т.е. не важно - какое его состояние было в прошлом. Такие процессы называются процессами без памяти (условие марковости процесса функционирования СМО рассмотрим ниже).

*) А.А. Марков (1856-1922) – профессор Петербургского университета, ученик П.Л.Чебышева, один из крупнейших представителей петербургской школы теории вероятностей. Основные работы Маркова по теории вероятностей посвящены предельным теоремам для сумм независимых величин, в том числе связанных в цепь.

Если свойство марковости процесса нарушается, то математическое его описание существенно усложняется и получить его аналитическое описание удается лишь в редкий случаях, зачастую достаточно приближенно. Тогда для определения характеристик СМО (решение прямой задачи) широко используется построение имитационных моделей.

Решение обратных-оптимизационных задач ИО для СМО, т.е. поиск наилучших параметров системы – числа каналов обслуживания, их производительности, дисциплины обслуживания и т.п. – с таким уровнем выбранного критерия эффективности системы в целом рассматривается обычно как отдельная задача. В ней модели ТМО используются для определения значений критерия эффективности системы или заданного значения ее параметров.

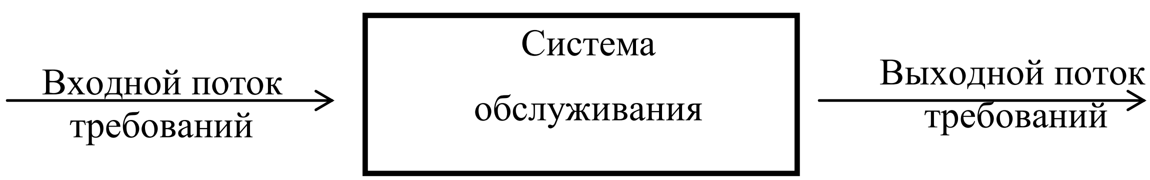
Оптимизация работы СМО может производиться под разными углами зрения: с точки зрения ее «владельца» и с точки зрения «клиентов». При этом перекос в ту или иную сторону может привести к потерям для другой стороны.

Поэтому при выборе критерия эффективности необходимо учитывать интересы обеих сторон и задача оптимизации СМО часто решается как многокритериальная.

Лекция №4.

Структура СМО и ее основные характеристики

В общем виде СМО можно представить как:



Входной поток требований–случайная последовательность требований, поступающих на обслуживание. Его основные характеристики:

- распределение числа требований в единицу времени;
- распределение времени между двумя соседними требованиями;
- ограниченность потока (ограничен, неограничен).

Система обслуживания характеризуется механизмом обслуживания и дисциплиной обслуживания.

Механизм обслуживания задается:

- распределением времени обслуживания;
- пропускной способностью (число каналов обслуживания);
- доступностью для обслуживания (полнодоступные – возможны всегда и неполнодоступные – обслуживание лишь в определенные периоды).

Дисциплина обслуживания – регламентирует порядок поступления требований на обслуживание из очереди:

- без приоритетов – все время в одинаковом порядке (первый пришел – первый обслужился, последний пришел – первый обслужился и т.д.).

- с приоритетами, - когда порядок обслуживания определяется не местом в очереди, а рангом требования. При этом приоритеты бывают:

- абсолютные – требование более высокого ранга исключает из идущего обслуживания требования меньшего ранга;
- относительные- начинает обслуживаться раньше требований низкого ранга.

По способу образования очереди СМО делятся на системы с потерями и без потерь.

Системы с потерями (отказами) или системы без ожидания. В них требование, поступившее в систему в момент, когда все каналы заняты, покидает систему.

Эффективность функционирования таких систем характеризуется:

- -вероятностью потери требования (отказа);
- -среднее число отказов за заданный временной интервал;
- -среднее число занятых каналов;

Системы без потерь (отказов) или системы с ожиданием. В них, в случае, если все каналы заняты, требования становятся в очередь на обслуживание. Длина очереди и время ожидания не ограничены.

Такие системы характеризуются:

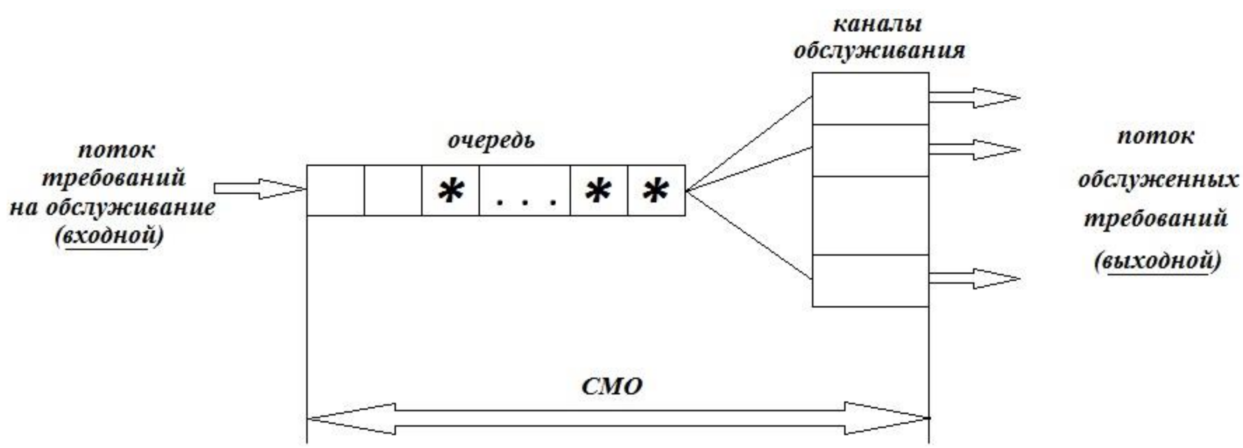
- -средней длиной очереди;
- -средним временем ожидания обслуживания;
- -средним числом занятых каналов.

Смешанные системы. В них накладываются ограничения на длину очереди, время пребывания в очереди, время обслуживания. Для них дополнительно могут использоваться следующие характеристики:

- -среднее число требований, покинувших систему до начала обслуживания;
- -среднее число требований, покинувших систему во время обслуживания;
- -время обслуживания требований, покинувших систему во время обслуживания.

Выходной поток требований – образуется требованиями, покидающими систему как обслуженными, так и не обслуженными из очереди или каналов обслуживания. Описывается теми же статистическими характеристиками, что и входной поток.

Схематически СМО можно представить в виде



Для удобства в 1953 году **Д.Г. Кендалл** предложил классифицировать СМО по совокупности трех основных признаков для подобных СМО. А именно

a/b/n,

где:

а- характер входящего потока, а конкретнее закон распределение интервала между требованиями;

b- закон распределения времени обслуживания одного требования;

n- число одновременно обслуживаемых требований (число каналов обслуживания СМО).

Для обозначения вида закона распределения входного потока**(а)** и потока обслуживания **(b**) обычно используют следующие обозначения:

- М – марковское – экспоненциальный закон распределения интервалов времени между требованиями во входном потоке и в потоке обслуживания;

- D – детерминированные интервалы времени между требованиями;

- $E_к$ - распределение Эрланга К-ого порядка;

- N – нормальное распределение;

- G – произвольный закон распределения.

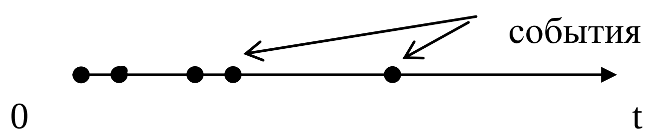
Кроме этого часто дополнительно указывают **m -** максимальную длину очереди (максимальное число требований в системе);

Так, <u>например,</u> **M/M/3 /m=0** обозначает трехканальную СМО без очереди с простейшим потоком на входе и экспоненциальным распределением времени обслуживания, а **E/D/1/m** обозначает одноканальную СМО с Эрланговским потоком требований на входе, постоянным временем обслуживания и ограниченной очередью.

Одним из важнейших факторов, определяющих возможность нахождения основных характеристик СМО аналитически, является <u>вид потоков</u> требований (заявок) на входе системы и при их обслуживании. Рассмотрим их.

<u>Входной поток требований</u>

<u>Потоком событий</u> (требований) называется последовательность одинаковых по своему содержанию событий (явлений), происходящих через какие-то промежутки времени.



Если события, образующие поток событий, одинаковы по своей природе, то *<u>поток однородный</u>*.

<u>Регулярный (детерминированный) поток</u>. В нем требования поступают в систему через определенные (как правило одинаковые) неслучайные моменты времени по одному (в общем случае – пакетами). <u>Интенсивность потока</u> – число требований, поступающих в единицу времени будет

$\lambda = \frac{1}{a}$**,** где **a** – интервал времени между двумя соседними требованиями.

<u>Примерами</u> такого потока служат: движение поездов через станцию строго по расписанию, поток деталей на конвейере, сеансы в кинотеатрах, занятия в институте…

СМО с детерминированным потоком рассчитываются достаточно просто. Однако в случае нарушения детерминированности результаты расчета оказываются непригодными.

Простейший Пуассоновский поток

Это наиболее распространенная форма описания входного потока, позволяющая достаточно просто исследовать СМО аналитически. В нем события происходят в неопределенные (случайные) моменты времени. Промежутки времени между событиями также являются случайными.

Пуассоновский поток можно определить как **поток, в котором число требований, поступающих в заданном временном интервале времени τ, распределено по закону Пуассона** с параметром **a=λ·τ**, где **λ** - интенсивность потока, т.е. среднее число требований, поступающих в единицу времени, а вероятность, что на интервале **τ** возникнет ровно k требований, равна

$$\mathbf{P_k(\tau) = \frac{(\lambda\tau)^k}{k!} \cdot e^{-\lambda\tau}, k = 0, 1, 2 \ldots}$$

Согласно свойству закона Пуассона величина $\mathbf{a = \lambda \cdot \tau}$ есть математическое ожидание (среднее значение) и дисперсия числа требований, поступающих на интервале времени τ.

Величина **λ** - (***интенсивность потока требований***) - одна из основных характеристик потока равна математическому ожиданию числа требований, поступающих в единицу времени.

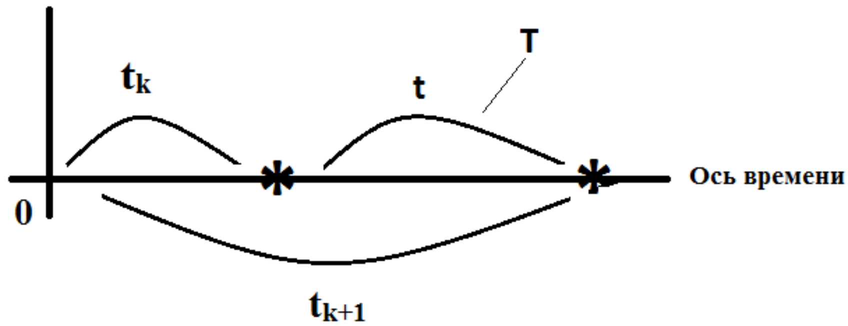
Простейший Пуассоновский поток обладает свойствами стационарности, ординарности и отсутствия последействия.

Стационарность означает, что статистические характеристики потока (распределение числа требований в единицу времени, распределение интервала между соседними требованиями) во времени не меняются.

Ординарность означает практическую невозможность одновременного появления двух и более требований. Т.е. вероятность появления двух и более требований на интервале $\Delta t \to 0$ бесконечно мала по сравнению с вероятностью появления одного требования

Отсутствие последействия означает, что вероятность появления **k** требований на интервале [t, t+τ] не зависит от числа требований, появившихся до момента t и положения интервала τ на временной оси.

Найдем закон распределения интервала времени между двумя соседними требованиями для простейшего Пуассоновского потока.



Пусть T – случайная величина – интервал между соседними требованиями.

Вероятность того, что величина T примет значение большее или равное t, равна вероятности того, что на интервале продолжительностью t не возникнет ни одного требования, т.е.

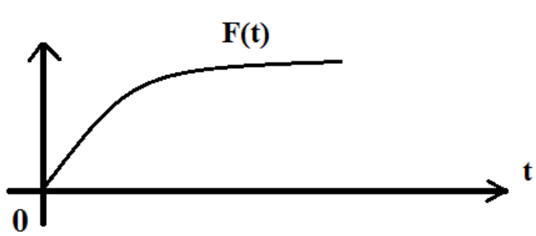
$$\mathbf{P(T \geq t) = P_0(t) = \frac{a^0}{0!} \cdot e^{-a} = e^{-\lambda\tau}},\text{ где } \underline{\mathbf{a = \lambda \cdot \tau}}.$$

Отсюда функция распределения будет (см. рис.)

$$\boldsymbol{F(t) = P(\mathrm{T} < t) = 1 - \mathrm{e}^{-\lambda\tau}} \qquad \text{при } \underline{t \geq 0}.$$

Функция плотности распределения равна

$$\mathbf{f(t) = \frac{d}{dt}F(t) = \lambda e^{-\lambda\tau}} \qquad \text{при } t \geq 0.$$



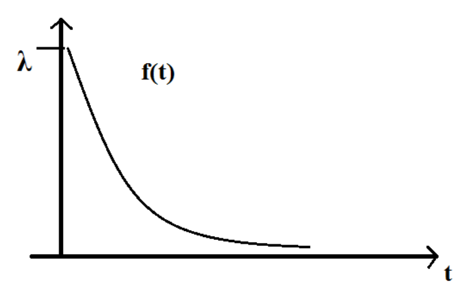


Рис. Вид законов распределения **F(t) и f(t)**.

Как видим это ***экспоненциальный закон распределения***. Его основные числовые характеристики:

$$\mathbf{m_t = \int_{-\infty}^{\infty} t \cdot f(t)dt = \frac{1}{\lambda}}, \quad \mathbf{D_t = \int_{-\infty}^{\infty} (t - m_t)^2 \cdot f(t)dt = \frac{1}{\lambda^2}}$$

Как видно из выражения для функции распределения, вид закона распределения и его параметры не зависят от положения интервала между требованиями на временной оси (стационарность) и от того, появлялись или нет требования раньше (отсутствие последействия). Т.е. если промежуток времени T, распределенного по экспоненциальному закону уже длился некоторое время τ < T, то это никак не влияет на закон распределения оставшейся части промежутка - он будет таким же, как и закон распределения всего промежутка T. Этим свойством обладает только экспоненциальный закон. Для него можно показать, также, и выполнение условия ординарности.

При анализе реальных входных потоков оказывается, что простейший поток имеет широкое распространение на практике. Это обуславливается предельными свойствами, согласно которым простейший поток возникает при суммировании большого числа взаимно независимых потоков малой интенсивности, обладающих свойствами стационарности и ординарности, при этом свойство отсутствия последействия для них (по отдельности) может и не выполняться.

Лекция № 5.

Другие виды входных потоков.

Нестационарный Пуассоновский поток. Если поток событий ординарный, без последействия, но с переменной плотностью числа событий, приходящихся на единицу времени $\lambda(t) \neq const$, то такой поток событий называется нестационарным Пуассоновским.

Для него число требований, поступающих на участке времени длиной τ, начинающемся в точке t_0 так же подчинено закону Пуассона

$$P(\tau/t_0) = \frac{a^k}{k!} \cdot e^{-a},$$

но значение параметра **a** зависит и от **τ** и от $\mathbf{t_0}$

$$a = \int_{t_0}^{t_0+\tau} \lambda(t)dt = \lambda_{cp} \cdot \tau = \tau \cdot \underbrace{\left(\frac{1}{\tau} \int_{t_0}^{t_0+\tau} \lambda(t)dt\right) =}_{\lambda_{cp}} \boldsymbol{\lambda_{cp}\, \tau}$$

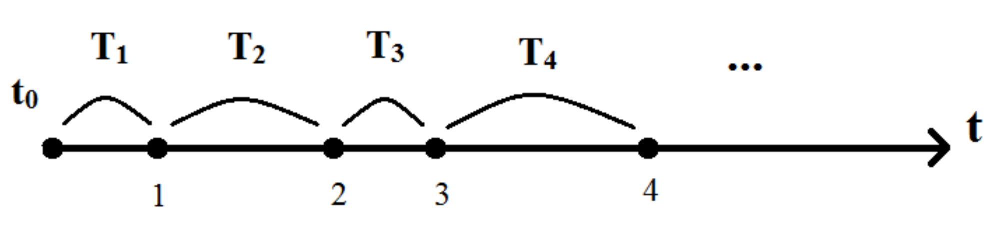
τ

Как видим, закон распределения интервала времени между двумя соседними требованиями отличается от экспоненциального, и зависит от вида функции **λ(t).**

Неординарный Пуассоновский поток. Сохраним стационарность и отсутствие последействия, но откажемся от ординарности следующим образом: Моменты прихода требований составляют простейший Пуассоновский поток, но в каждый из этих моментов $\mathbf{t_i}$ может прийти группа из **m** требований, причем **m** случайно и задается распределением **P(m)** – вероятностью того, что в группе будет **m** требований. Тогда вероятность прихода в случайный момент времени $\mathbf{(t_i)}$ **j** требований будет равна:

$$P_{i,j}(\Delta t) = \lambda \Delta t \cdot P(i, j)$$

Поток с ограниченным последействием. Рассмотрим ординарный поток событий.



Здесь 1,2,3,… события, T_1, T_2,… интервалы времени между событиями.

Если интервалы времени между последовательными событиями представляют собой независимые случайные величины, то такой поток событий называется потоком с ограниченным последействием (поток Пальма).

Потоком Пальма является поток отказов элементов в техническом устройстве, если отказы наступают независимо, а отказавший элемент мгновенно заменяется новым. Поток отказов часто называют потоком восстановлений т.к. в момент отказа происходит мгновенное восстановление элемента. Потоком с ограниченным последствием является, также, поток не обслуженных заявок, покидающих систему (*Теорема Пальма*: если на вход СМО поступает поток заявок типа Пальма, и в случае занятости каналов обслуживания заявка покидает систему не обслуженной (получает отказ), а время обслуживания заявок в системе подчинено экспоненциальному закону, то поток необслуженных заявок также является потоком Пальма.).

Частным случаем потока Пальма является простейший поток.

Другим частным случаем потоков Пальма являются потоки Эрланга, которые образуются «просеиванием» простейшего потока.

Поток Эрланга: если из простейшего потока будем удалять требования через одно, то оставшиеся требования образуют поток Эрланга первого порядка, если будем удалять требования парами, оставляя каждое третье, то они образуют поток Эрланга второго порядка и т.п. Если из простейшего потока будем удалять требования группами по **k** штук в каждой группе, то оставшиеся требования образуют поток Эрланга **k** –го порядка (см. рис).

```
|_*____*____*____*____*____*____*____*____*_____*_ Простейший поток
|_*_________*_________*_________*_________*_______ Поток Эрланга 1-го порядка
|_*______________*______________*_______________*_ Поток Эрланга 2-го порядка
|_*____________________*____________________*_______ Поток Эрланга 3-го порядка
```

Рис. Потоки Эрланга различного порядка.

Интервал времени между требованиями в потоке Эрланга К-го порядка подчинён закону распределения Эрланга К-го порядка как закон распределения суммы (К+1) независимых случайных величин с экспоненциальным законом распределения

$f_к(t) = \lambda * \frac{(\lambda t)^к}{к!} * e^{-\lambda t}$ при $t \geq 0$ с параметрами:

$m_к = (к+1) * m_0 = (к+1) * \frac{1}{\lambda}$ - мат. ожидание;

$D_К = (К+1) * D_0 = (К+1) * \frac{1}{\lambda^2}$ - дисперсия,

где: m_0, D_0 - параметры экспоненциального закона (закон Эрланг нулевого порядка).

Можно показать, что при увеличении порядка (К) поток Эрланга становится все более регулярным.

Кроме рассмотренных выше, на практике имеют место и другие входные потоки:

- Неординарные потоки (требования поступают не по одному, а группами);
- Сложные детерминированные потоки – неслучайные по сути, но со сложной функцией моментов поступления.
- Дискретные потоки. В них требования могут поступать только в определённые моменты времени. Например, на транспорте – по расписанию и т.д.

<u>Выходной поток. Время обслуживания требований</u>

Эффективность функционирования системы обслуживания, как уже говорилось ранее, определяется в частности временем обслуживания каждого требования. Длительность обслуживания в общем случае является случайной величиной, подчинённой некоторому закону распределения. В ТМО наиболее распространены следующие виды законов.

<u>1. Показательное распределение</u> длительности времени обслуживания. В этом случае время нахождения требования в канале обслуживания подчинено экспоненциальному закону распределения с параметром μ – среднее число заявок, обслуженных в канале за единицу времени (интенсивность обслуживания)

$g(t) = \mu * e^{-\mu t}$ при $t \geq 0$.

Математическое ожидание и дисперсия времени обслуживания заявки в канале будут:

$m_{(t)} = \frac{1}{\mu}$; $D_{(t)} = \frac{1}{\mu^2}$.

Показательное распределение времени обслуживания целесообразно применять в тех случаях, когда большее число требований требует для своего обслуживания сравнительно малого времени, а меньшее число требований – большого (<u>например,</u> на почте – много мелких отправлений (заявок) – марки, конверты,…. реже заказные письма, телеграммы). Совсем редко = отправление одним клиентом сразу большого числа заказной корреспонденции.

Основное достоинство экспоненциального распределения в задачах анализа СМО состоит в том, что оно не обладает памятью (не зависит от начала отсчета), и, следовательно, открывает возможности для построения марковской модели СМО.

<u>2. Эрланговское распределение</u> длительности обслуживания. Используется при анализе СМО в тех случаях, когда процесс обслуживания можно представить как К последовательных, независимых этапов, длительность обслуживания на каждом из которых подчинена экспоненциальному закону распределения с параметром μ (распределение Эрланга (К-1)-го порядка).

3. Нормальное распределение. Часто встречаются ситуации, когда время обслуживания хорошо сгруппировано вокруг среднего значения и отклонения от него распределены симметрично с экспоненциально убывающей плотностью на ветвях. В таких случаях напрашивается использование нормального распределения. Однако, применять его надо с осторожностью. Дело в том, что генерация значений времени обслуживания в соответствии с нормальным законом чревата тем, что возможно появление отрицательных чисел, чего ни в коем случае быть не должно. Отрицательных чисел практически не будет, если среднее время обслуживания в 5-6 раз больше, чем С.К.О.

4. Постоянная длительность обслуживания. Применяется для приближённого расчёта СМО с сильной идеализацией процессов обслуживания.

Кроме этих законов распределения могут применяться и *другие*:

- с нестационарной интенсивностью обслуживания
- с неординарным потоком заявок.
- с зависимостью времени обслуживания от типа требования и др.

В этих случаях получить простые аналитические выражения для анализа СМО часто не представляется возможным. Поэтому:

- либо используют приближённые вычисления, переходя к экспоненциальным законам с эквивалентными характеристиками;

- либо приводят немарковские модели к марковским, используя специальные приемы;

- либо используют для вычислений специальные модели, например имитационные.

Марковские системы массового обслуживания.

Под **случайным процессом** понимают изменение во времени состояний некоторой физической системы заранее неизвестным случайным образом. При этом под физической системой будем понимать любое техническое устройство, группу устройств, предприятие, отрасль, биологическую систему и т.д.

Случайный процесс, протекающий в системе называется **Марковским** – если для любого момента времени t_0 ,вероятностные характеристики процесса *в будущем* (t > t_0) зависят только от его состояния в данный момент времени t_0 (*в настоящем*) и не зависят от того, когда и как система пришла в это состояние (*в прошлом)*. Например, счетчик Гейгера, регистрирующий число космических частиц.

Марковские процессы принято делить на 3 вида:

1. *Марковская цепь* – процесс, состояния которого дискретны (т.е. их можно перенумеровать) и время, по которому он рассматривается, также дискретно (т.е. процесс может менять свои состояния только в определенные моменты времени). Такой процесс изменяется по шагам (иначе - по тактам).

2. *Дискретный марковский процесс* – множество состояний дискретно (можно перечислить), а время непрерывно (переход из одного состояния в другое – в любой момент времени).

3. *Непрерывный марковский процесс* – множество состояний и время - непрерывные.

На практике Марковские процессы в чистом виде встречаются не часто. Однако нередко приходится иметь место с процессами, для которых влиянием предыстории можно пренебречь. Кроме того, если все параметры процесса из «прошлого», от которых зависит «будущее» включить в состоянии системы в «настоящем», то ее также можно рассматривать как Марковскую. Однако это часто приводит к значительному росту числа учитываемых переменных и невозможности получить решение задачи.

В исследование операций большое значение занимают так называемые *Марковские случайные процессы с дискретными состояниями и непрерывным временем*.

Процесс называется процессом с дискретными состояниями, если все его возможные состояния S_0, S_1,... можно заранее перечислить (перенумеровать). Переход системы из состояния в состояние переходит практически мгновенно – скачком.

Процесс называется процессом с непрерывным временем, если моменты перехода из состояния в состояние могут принимать любые случайные значения на временной оси.

Например: Техническое устройство **S состоит из двух узлов**, каждый из которых в случайный момент времени может выйти из строя (отказать). После этого мгновенно начинается ремонт узла (восстановление),который продолжается случайное время.

Возможны следующие состояния системы :

S_0- оба узла исправны;

S_1 - первый узел ремонтируется , второй исправен.

S_2 – второй узел ремонтируется , первый исправен

S_3- оба узла ремонтируются .

Переход системы из состояния в состояние происходит в случайные моменты времени практически мгновенно. Состояния системы и связь между ними удобно отобразить с помощью **графа состояний** (см. рис.).

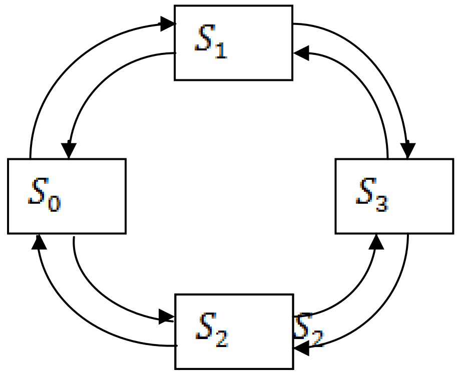
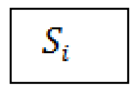


Рис. Граф состояний

Здесь:

 - переходы; 

 - состояния.

Переходы $S_0 \rightleftarrows S_3$ и $S_0 \rightleftarrows S_2$ отсутствуют, т.к. отказы и восстановления элементов происходят независимо и случайно и вероятность одновременного выхода из строя (восстановления) двух элементов бесконечно мала, и ею можно пренебречь.

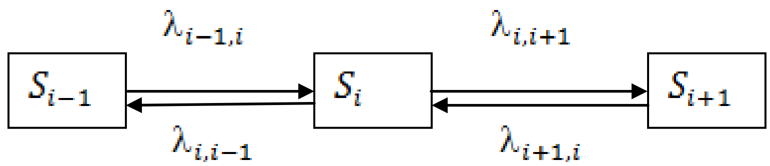
Если все потоки событий, переводящие систему **S** из состояния в состояние **простейшие**, то **процесс,** протекающий в такой системе, **будет Марковским**. Это обуславливается тем, что простейший поток не обладает последействием, т.е. в нем «будущее» не зависит от «прошлого» и, кроме того, он обладает свойством ординарности, так как вероятность одновременного появления двух и более событий бесконечно мала. Т.е. невозможен переход из состояния в состояние, минуя несколько промежуточных состояний.

Для наглядности на графе состояний удобно у каждой стрелки перехода проставить интенсивность того потока событий, который переводит систему из состояния в состояние по данной стрелке ($\boldsymbol{\lambda_{ij}}$ -интенсивность потока событий, переводящего систему из состояния $\boldsymbol{S_i}$ **в** $\boldsymbol{S_j}$). Такой граф называется **размеченным**. В нашем случае это будут переходы: $\boldsymbol{\lambda_{01}}$, $\boldsymbol{\lambda_{10}}$, $\boldsymbol{\lambda_{02}}$, $\boldsymbol{\lambda_{20}}$, $\boldsymbol{\lambda_{13}}$, $\boldsymbol{\lambda_{31}}$, $\boldsymbol{\lambda_{23}}$, $\boldsymbol{\lambda_{32}}$ **.**

Используя размеченный граф состояний системы можно построить математическую модель данного процесса.

Пусть система в момент времени **t** находится в состоянии $\boldsymbol{S_i}$.

Рассмотрим переходы системы из некоторого состояния S_i в предыдущее S_{i-1} или последующее S_{i+1}. Фрагмент графа состояний в этом случае будет выглядеть следующим образом:



Лекция № 6.

Обозначим P_i(t)- вероятность i-ого состояния системы – вероятность того, что система в момент времени **t** находится в состоянии S_i. Для любого момента времени t справедливо $\sum_i P_i(t)=1$.

Определим вероятность того, что и в момент времени **t+Δt** система будет находиться в состоянии $\boldsymbol{S_i}$. Это может быть в следующих случаях:

Система находилась в состоянии S_i и за время Δt из него не вышла.

Система находилась в состоянии S_{i-1} и за время Δt перешла в состояние S_i.

Система находилась в состоянии S_{i+1} и за время Δt перешла в состояние S_i.

Можно показать, что при простейшем потоке требований (событий), переводящих систему из состояния в состояние, (экспоненциальный закон распределения интервала времени между требованиями с интенсивностью **λ)**, вероятность того, что за время **Δt** (при **$\Delta t \to 0$**) не возникнет ни одного требования (с точностью до бесконечно малой **(б.м.)** величины) будет равна

P (k=0/$\Delta t \to 0$) = (1- $\lambda * \Delta t$) + (б.м.).

Вероятность того, что за время **Δt** (при **$\Delta t \to 0$**) возникнет **хотя бы одно** требование, будет равна

P(k$\geq$0/$\Delta t \to 0$) = 1 - P(k=0) = 1 – (1 - $\lambda * \Delta t$) = $\lambda * \Delta t$ + (б.м.).

Учитывая свойство ординарности потока (**P(k$\geq$2/$\Delta t \to 0$) =0)** получим вероятность появления **ровно одного** требования

P (k=1/$\Delta t \to 0$) = $\lambda * \Delta t$ + (б.м.).

Тогда вероятности соответствующих переходов из состояния в состояние, с учетом общей интенсивности таких переходов будут равны:

1) ***Система находилась в состоянии S_i и за время Δt из него не вышла.*** То есть за это время возникло **k = 0** требований с суммарной интенсивностью **($\lambda_{i,i-1} + \lambda_{i,i+1}$)**

P (S_i/S_i)=1 – ($\lambda_{i,i-1}+\lambda_{i,i+1}$) $*$ Δt + **(б.м.).**

2) ***Система находилась в состоянии S_{i-1} и за время Δt перешла в состояние S_i*** (возникло ровно **k = 1** требование с интенсивностью **$\lambda_{i-1,i}$).**

P (S_i/S_{i-1}) = $\lambda_{i-1,i} * \Delta t$ + (б.м.).

3) ***Система находилась в состоянии S_{i+1} и за время Δt перешла в состояние S_i*** (возникло ровно **k = 1** требование с интенсивностью **$\lambda_{i+1,i}$).**

P (S_i/S_{i+1}) = $\lambda_{i+1,I} * \Delta t$ + (б.м.).

Тогда ***вероятность, что система в момент времени (t+Δt) будет в состоянии S_i*** равна

$$\mathbf{P_i(t+\Delta t)=P_i(t) * P(S_i/S_i)+P_{i-1}(t) * P(S_i/S_{i-1})+P_{i+1}(t) * P(S_i/S_{i+1}) =}$$

$$\mathbf{= P(S_i/S_{i+1}) = P_i\,(t) * [1 - (\lambda_{i,i-1} + \lambda_{i,i+1})] + P_{i-1}(t) * \lambda_{i-1,i} * \Delta t + P_{i+1}(t) * \lambda_{i+1,I} * \Delta t}$$

+ (б.м.).

Вычтем из обеих частей $P_i(t)$, разделим на Δt и, перейдя к пределу при $\Delta t \to 0$, получим

$$\frac{\mathbf{dP_i(t)}}{\mathbf{dt}} = -\mathbf{P_i(t)\cdot\left(\lambda_{i,i-1}+\lambda_{i,i+1}\right)+P_{i-1}(t)\cdot\lambda_{i-1,i}+P_{i+1}(t)\cdot\lambda_{i+1,i}}$$

Данные уравнения называются уравнениями **Колмогорова-Чепмена для дискретного марковского процесса**.

Подставив соответствующие значения интенсивностей переходов из состояний в состояния, получим систему дифференциальных уравнений, описывающих изменение вероятностей состояний рассматриваемой системы как функций времени.

$$\frac{dP_0(t)}{dt} = -(\lambda_{01}+\lambda_{02})\cdot P_0(t) + \lambda_{10}\cdot P_1(t) + \lambda_{20}\cdot P_2(t)$$

$$\frac{dP_1(t)}{dt} = -(\lambda_{10}+\lambda_{13})\cdot P_1 + \lambda_{01}P_0 + \lambda_{31}P_3$$

$$\frac{dP_2(t)}{dt} = -(\lambda_{20}+\lambda_{23})\cdot P_2 + \lambda_{02}\cdot P_0 + \lambda_{32}\cdot P_3$$

$$\frac{dP_3(t)}{dt} = -(\lambda_{31}+\lambda_{32})\cdot P_3 + \lambda_{13}\cdot P_1 + \lambda_{23}\cdot P_2$$

Задав начальные условия (например, $\mathbf{P_0(t=0)=1,\ P_i(t=0)=0\quad i\neq 0}$) и решив их, получим выражения для вероятностей состояния системы $\mathbf{P_i(t)}$ как функций времени. Аналитические решения достаточно просто получить, если число уравнений $\leq 2,3$. Если их больше, то обычно решают уравнения численно на ЭВМ (например, методом Рунге-Кутта).

В теории случайных процессов доказано, что если число состояний системы *конечно*, и из каждого из них можно за конечное число шагов перейти в любое другое, то *существует предел* к которому стремятся вероятности $P_i(t)$ при $\mathbf{t\to\infty}$**.** Такие вероятности называются ***финальными вероятностями*** состояний, а установившийся режим - ***стационарным режимом*** функционирования системы.

Имея графики изменения всех $\mathbf{P_i(t)}$ можно найти ***время переходного процесса*** (время достижения значений финальных вероятностей с заданной точностью – время вхождения в **δ%** «трубку»).

Так как в стационарном режиме все $P_i(t) = const$, следовательно, все $\frac{dP_i(t)}{dt}$=0. Приравняв в системе уравнений левые части нулю и, дополнив их уравнением $\sum_i P_i$=1, получим систему линейных алгебраических уравнений, решив которую найдём значения финальных вероятностей.

Пример. Пусть в нашей системе интенсивности отказов и восстановления элементов следующие

Отказы: 1эл: $\lambda_{отк} = \lambda_{01} = \lambda_{23} = 1$;

2эл: $\lambda_{отк} = \lambda_{02} = \lambda_{13} = 2$.

Ремонт: 1эл: $\mu_{рем1} = \lambda_{10} = \lambda_{32} = 2$;

2эл: $\mu_{рем2} = \lambda_{20} = \lambda_{31} = 3$.

Подставив соответствующие значения интенсивностей в уравнения, получим:

$$\left.\begin{array}{l} P_0+P_1+P_2+P_3=1 \\ 0= -(1+2)\,P_0+2P_1+3\,P_2 \\ 0= -(2+2)\,P_1+1P_0+3P_3 \\ 0= -(1+3)\,P_2+2P_0+2P_3 \\ 0= -(2+3)\,P_3+2P_1+1P_2 \end{array}\right\}$$

Решив данную систему, найдём значения финальных вероятностей:

$P_0=6/15=0.4$; $P_1=3/15=0.2$; $P_2=4/15=0.27$; $P_3=2/15\approx0.13$.

Т.е. в стационарном состоянии система в среднем:

- **40%** находится в состоянии S_0 (оба узла исправны);
- **20%**- в состоянии S_1 (1-й эл-т ремонтируется, 2-й исправен);
- **27%**- в состоянии S_2 (2-й эл-т ремонтируется, 1-й исправен);
- **13%**- в состоянии S_3 – оба эл-та в ремонте.

Знание финальных вероятностей позволяет оценить среднюю эффективность работы системы и загрузку службы ремонта (прямая задача ИО).

Пусть система в единицу времени приносит **доход**: в состоянии S_0 - 8 усл.ед.; в состоянии S_1 - 3 усл.ед.; в состоянии S_2 - 5 усл.ед.; в состоянии S_3 - =0.

Стоимость **ремонта** в единицу времени для 1-го эл-та составляет 1 усл.ед. в единицу времени (состояния S_1, S_3), 2-го эл-та - 2 усл.ед. (состояния S_2, S_3) Тогда в стационарном режиме:

Доход системы (математическое ожидание) в единицу времени будет:

$W_{дох}=8P_0+3P_1+5P_2+0P_3=8\cdot0.4+3\cdot0.2+5\cdot0.27+0\cdot0.13=5.15$ усл.ед.

Стоимость ремонта в ед. времени:

$W_{рем}=0P_0+1P_1+2P_2+(1+2)P_3=0\cdot0.4+1\cdot0.2+2\cdot0.27+3\cdot0.13=1.39$ усл.ед.

Прибыль (математическое ожидание) в единицу времени

$\mathbf{W = W_{дох} - W_{рем} = 5.15 - 1.39 = \underline{\underline{3.76}}}$ **усл.ед.**

Проведя определённые расходы по модернизации системы можно изменить интенсивности $\boldsymbol{\lambda}$ и $\boldsymbol{\mu}$ и, соответственно, эффективность системы в целом. Целесообразность таких расходов можно оценить, проведя пересчёт $\mathbf{P_i}$. и соответствующих показателей эффективности системы **W**. (обратная задача ИО).

Модели марковских СМО.

Как видно из предыдущего в случае, если для системы **S** имеется размеченный граф состояний и она является марковской, то для нее легко могут быть составлены уравнения Колмогорова - Чепмена, на основании которых рассчитывают значения вероятностей состояния системы как в произвольные моменты времени, так и в стационарном режиме (если он существует).

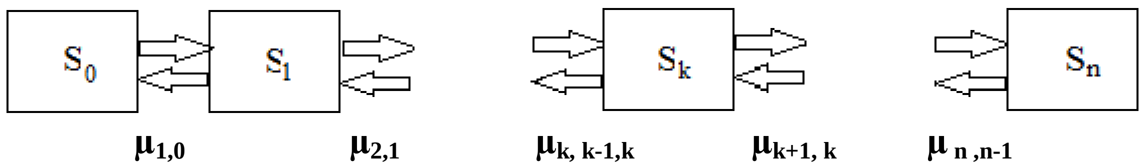
Общий интерес имеет случай, когда граф состояний системы представляет собой так называемую «**схему размножения и гибели**». В этом случае для финальных вероятностей удается получить достаточно простые выражения.

Особенность этого состоит в том, что все состояния системы в графе состояний можно вытянуть в одну цепочку, в которой каждое из средних состояний системы от S_1 до S_{n-1} связано с соседними состояниями, справа и слева, прямой и обратной стрелкой (переходом), а крайние состояния S_0 и S_n только с одним соседним (внутренним) состоянием.

Термин «*схема размножения и гибели*» берет начало от биологических задач, где подобной схемой описывается изменение численности популяции. Процесс размножения и гибели соответствует как бы двум потокам: «*размножения*» - переход от состояния S_{k-1} к состоянию S_k с интенсивностью $\lambda_{k-1,k}$ за счет возникновения требований в системе и «*гибели*» - переход от S_{k+1} к S_k с интенсивностью $\mu_{k+1,k}$ – за счет уменьшения («обслуживания») числа требований.

Для схемы размножения и гибели граф состояний системы имеет вид:

$\lambda_{0,1}$ $\lambda_{1,2}$ $\lambda_{k-1,k}$ $\lambda_{k,k+1}$ $\lambda_{n-1,n}$



В случае, если в системе существуют только потоки требований «вперед», т.е. за счет их возникновения, то процесс рассматривается как *«процесс чистого размножения»*. Если изменение состояний системы происходит только «назад», за счет обслуживания требований, то говорят о *«процессе чистой гибели»*.

Пусть все потоки событий, переводящие систему S из состояния в состояние – простейшие. Тогда уравнения Колмогорова-Чепмена для неё могут быть записаны в виде:

$$\frac{dP_0}{dt} = -\lambda_{01}P_0 + \mu_{10}P_1$$

$$\frac{dP_1}{dt} = -(\lambda_{12} + \mu_{10})P_1 + \lambda_{01}P_0 + \mu_{21}P_2$$

……………………..

$$\frac{dP_k}{dt} = -(\lambda_{k,k+1} + \mu_{k,k-1})P_k + \lambda_{k-1,k}P_{k-1} + \mu_{k+1,k}P_{k+1}$$

……………………...

$$\frac{dP_n}{dt} = -\mu_{n,n-1}P_n + \lambda_{n-1,n}P_{n-1}$$

Для стационарного режима (при $\mathbf{t} \to \infty$; $\frac{\mathrm{dP_k}}{\mathrm{dt}} = \mathbf{0}$, при $\boldsymbol{k} = \mathbf{0}, \boldsymbol{n}$) их можно записать в следующем виде:

$$\lambda_{01}P_0 = \mu_{10}P_1\ ;$$

$$(\lambda_{12} + \mu_{10})P_1 = \lambda_{01}P_0 + \mu_{21}P_2;$$

……………………...

$$(\lambda_{k,k+1} + \mu_{k,k-1})P_k = \lambda_{k-1,k}P_{k-1} + \mu_{k+1,k}P_{k+1};$$

……………………..

$$\mu_{n,n-1}P_n = \lambda_{n-1,n}P_{n-1}.$$

Подставив предыдущее уравнение в последующее получим (например, для второго уравнения):

$$\lambda_{12}P_1 + \mu_{10}P_1 = \lambda_{01}P_0 + \mu_{21}P_2\ .$$

Подставив из первого уравнения значение $\lambda_{01}P_0 = \mu_{10}P_1$ получим

$$\lambda_{12}P_1 = \mu_{21}P_2\ .$$

Подставив, аналогично, каждое предыдущее уравнение в последующее получим следующую систему:

$$\lambda_{01}P_0 = \mu_{10}P_1$$

$$\lambda_{12}P_1 = \mu_{21}P_2$$

....

$$\lambda_{k-1,k}P_{k-1} = \mu_{k,k-1}P_k$$

....

$$\lambda_{n-1,n}P_{n-1} = \mu_{n,n-1}P_n$$

Каждое из уравнений этой системы характеризует баланс «<u>прямых</u>» и «<u>обратных</u>» потоков из состояния в состояние с учетом вероятностей состояний. Выразим из них вероятности всех последующих состояний через предыдущее:

$$P_1 = \frac{\lambda_{01}}{\mu_{10}}P_0$$

$$P_2 = \frac{\lambda_{12}}{\mu_{21}}P_1 = \frac{\lambda_{01}\lambda_{12}}{\mu_{10}\mu_{21}}P_0$$

…

$$P_k = \frac{\lambda_{k-1,k}}{\mu_{k,k-1}}P_{k-1} = \frac{\lambda_{01}\lambda_{12}\dots\lambda_{k-1,k}}{\mu_{10}\mu_{21}\dots\mu_{k,k-1}}P_0$$

…

$$P_n = \frac{\lambda_{n-1,n}}{\mu_{n,n-1}} P_{n-1} = \frac{\lambda_{01}\lambda_{12}\dots\lambda_{n-1,n}}{\mu_{10}\mu_{21}\dots\mu_{n,n-1}} P_0$$

Таким образом, все финальные вероятности состояний системы можно выразить через интенсивности потоков и вероятность начального состояния P_0. Дополнив систему нормирующим условием $\sum_{i=0}^{n} P_i = 1$ получим:

$$P_0 * (1 + \frac{\lambda_{01}}{\mu_{10}} + \frac{\lambda_{01}\lambda_{12}}{\mu_{10}\mu_{21}} + \dots + \frac{\lambda_{01}\lambda_{12}\dots\lambda_{n-1,n}}{\mu_{10}\mu_{21}\dots\mu_{n,n-1}}) = 1.$$

Отсюда получим выражение для P_0:

$$P_0 = \left(1 + \frac{\lambda_{0\,1}}{\mu_{1\,0}} + \frac{\lambda_{0\,1} * \lambda_{1\,2}}{\mu_{1\,0} * \mu_{2\,1}} + \dots + \frac{\lambda_{0\,1} * \lambda_{1\,2} * \dots * \lambda_{n-1,\;n}}{\mu_{1\,0} * \mu_{2\,1} * \dots * \mu_{n,\;n-1}}\right)^{-1}$$

Или …$P_0 = \left(1 + \sum_{k=1}^{n} \prod_{i=1}^{k} \frac{\lambda_{i-1,\;i}}{\mu_{i,\;i-1}}\right)^{-1}$.

Если *число состояний* ***n*** *конечно*, то ряд, стоящий в выражении для P_0 – сходится и, следовательно, вероятности P_0 и P_k ($1 \le k \le n$) отличны от нуля. В этом случае **в системе существует стационарный режим**.

Если же число возможных состояний **n** системы неограниченно (**n = ∞**), то ряд может расходиться. В этом случае **режима статического равновесия** в системе **не существует**. Тогда P_0 и остальные P_k ($k > 0$) будут равны нулю.

Для того, **чтобы режим статического равновесия существовал,** необходимо, чтобы ряд в выражении для P0 сходился. Для этого необходимо, чтобы, *начиная с некоторого состояния* $\boldsymbol{i_0}$,

выполнялось условие $\frac{\lambda_{i-1,\;i}}{\mu_{1,\;i-1}} < 1$ **для всех** $\boldsymbol{i > i_0}$.

т.е. интенсивность процесса «гибели» была выше, чем интенсивность процесса «размножения». В противном случае номер состояния **k** системы будет неуклонно возрастать.

Для стационарного режима вероятности $\mathbf{P_k}$ будут равны

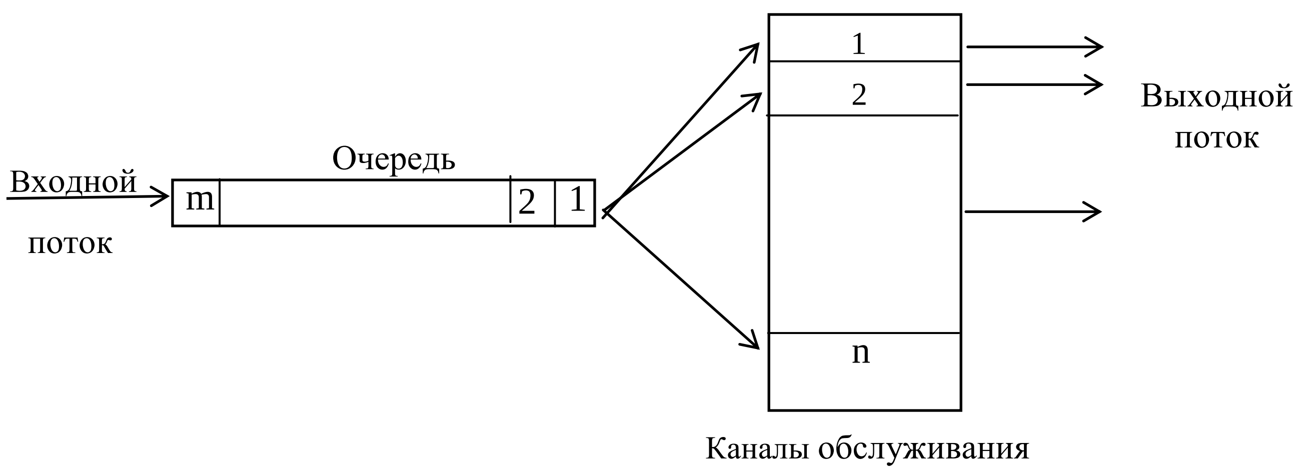
$$P_k = \prod_{i=1}^{k} \frac{\lambda_{i-1,\, i}}{\mu_{i,\, i-1}} * P_0 \qquad \underline{1 < k < n.}$$

Зная значения вероятностей $\mathbf{P_k}$ можно вычислить основные характеристики рассматриваемой СМО.

Лекция №7.

Определение основных характеристик СМО

Рассмотрим Систему Массового Обслуживания следующего вида



Здесь:**n –** число каналов обслуживания, **m** –максимальная длина очереди.

В любой момент времени **t** заявка из входного потока <u>либо</u> <u>становится</u> в один из каналов <u>на обслуживание</u>, <u>либо</u> <u>в очередь</u>, если все каналы заняты.

Под <u>**состоянием СМО**</u> будем понимать <u>**k- число заявок**</u>, находящихся в ней. Оно будет изменяться от нуля, когда заявок в системе нет, до **(n+m)**, когда все каналы и все места в очереди заняты. Подобную СМО можно представить в виде графа состояний, с помощью которого можно записать уравнения размножения и гибели, описывающие её поведение. Решив их для любого произвольного момента времени **t** (в частности для стационарного режима при **t** = ∞)

можно определить значения вероятностей нахождения СМО в каждом из состояний S_k

P_k (t)   k = 0, (n+m)   t > 0.

Зная их можно определить *основные характеристики СМО*, такие как:

1. ***Среднее число занятых каналов обслуживания***

$$m_{\text{кан}} = P_0 * 0 + P_1 * 1 + \dots + P_n * n + n * (P_{n+1} + \dots + P_{n+m})$$

2. ***Средняя длина очереди***

$$m_{\text{оч}} = 0 * (P_0 + \dots + P_n) + 1 * P_{n+1} + 2 * P_{n+2} + \dots + m * P_{n+m}$$

3. ***Вероятность, что требование, поступившее в момент времени t, получит отказ*** (т.к. в это время в системе вся очередь и все каналы обслуживания заняты):

$$P_{\text{отк}}(t) = P_{n+m}(t)$$

4. ***Вероятность, что поступившее требование будет обслужено***

$$P_{\text{обсл}} = 1 - P_{\text{отк}} = P_0 + \dots + P_n + \dots + P_{n+m-1} = 1 - P_{n+m}$$

Аналогично могут быть найдены:

5. ***Вероятность, что поступившее в систему требование сразу поступит в канал на обслуживание***

$$P_{\text{кан}} = P_0 + \dots + P_{n-1}$$

6. ***Вероятность, что поступившее требование попадёт в очередь***

$$P_{\text{оч}} = P_n + P_{n+1} + \dots + P_{n+m-1}$$

Формула Литтла

Среди различных математических выражений, позволяющих определять основные характеристики СМО **в стационарном режиме,** особое место занимает выражение, устанавливающее **связь между средним числом заявок,** находящихся **в системе** (обслуживающих или стоящих в очереди), **и средним временем пребывания заявки в системе.**

Рассмотрим любую СМО – одноканальную, многоканальную, марковскую, немарковскую, с ограниченной или неограниченной очередью и связанные с нею два потока: поток заявок, приходящих в систему и поток заявок, покидающих систему.

Если в системе установился стационарный режим, то среднее число заявок $\boldsymbol{\lambda_{сист}}$, входящих в СМО за единицу времени **,** равно среднему числу заявок $\boldsymbol{\mu_{сист}}$, покидающих её за это же время . То есть, оба потока имеют одинаковую интенсивность.

Можно показать, что для любой заявки, вошедшей в систему в некоторый момент времени t , время её нахождение в системе будет определятся числом заявок, находящихся в системе $m_{сист}$ и интенсивностью потока заявок $\boldsymbol{\mu_{сист}}$, покидающих систему за единицу времени, то есть

$$t_{сист} = m_{сист} * \frac{1}{\mu_{сист}}$$ **Формула Литтла.**

Так как в стационарном режиме всегда выполняется условие $\boldsymbol{\mu_{сист}=\lambda_{сист}}$, то обычно это выражение записывается в виде

$$t_{сист} = \frac{1}{\lambda_{сист}} * m_{сист} = \frac{1}{\mu_{сист}} * m_{сист}$$ **Формула Литтла**

Формула Литтла: для любой СМО, при любом характере потока заявок, при любом распределении времени обслуживания, при любой дисциплине обслуживания **среднее время пребывания заявки в системе $\boldsymbol{t_{сист}}$ равно среднему числу заявок в системе $\boldsymbol{m_{сист}}$, деленному на интенсивность потока заявок, поступивших в систему $\boldsymbol{\lambda_{сист}}$ (покинувших систему $\boldsymbol{\mu_{сист}}$) .**

С помощью формулы Литтла можно найти следующие величины.

Среднее время пребывания заявки в очереди:

$$t_{оч} = \frac{m_{оч}}{\lambda_{сист}}$$

Среднее время пребывания заявки в канале обслуживания:

$$t_{кан} = \frac{m_{кан}}{\lambda_{сист}} .$$

Так как известно, что среднее время пребывания требования в канале **t_{кан}** равно **1/μ ,** отсюда можно легко получить выражение для среднего числа занятых каналов:

$$t_{кан} = t_{кан} = \frac{m_{кан}}{\lambda_{сист}} = \frac{1}{\mu} \quad \text{отсюда} \qquad m_{кан} = \frac{\lambda сист}{\mu} .$$

Следует отметить, что во всех случаях $\boldsymbol{\lambda}_{сист}$ характеризует интенсивность требований, вошедших в систему, в не весь входной поток заявок. Если часть требований из потока заявок на входе теряется (например, в системах с отказами), то необходимо вместо интенсивности требований, подошедших к системе $\boldsymbol{\lambda}_{вх}$, взять интенсивность требований, вошедших в систему, т.е

$$\lambda_{сист} = (1 - P_{отк}) * \lambda_{вх} ,$$

где $(1 - P_{отк})$ - вероятность, что требование поступит в систему (не получит отказ). С учетом сказанного выражение для среднего числа занятых каналов примет вид

$$m_{кан} = \frac{\lambda сист}{\mu} = \frac{\lambda вх * (1 - Pотк)}{\mu} = \rho * (1 - Pотк) .$$

<u>Модели СМО без очереди. СМО без ожидания. Уравнения Эрланга.</u>

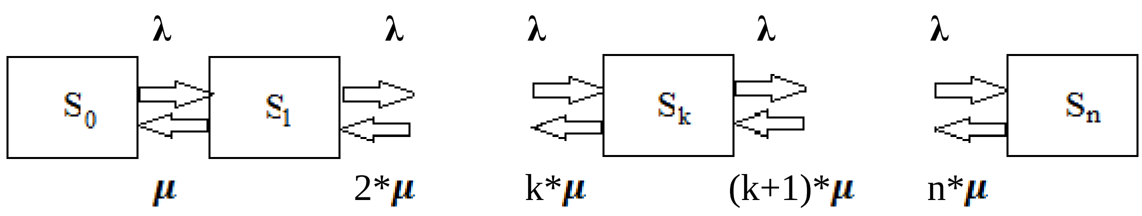
Рассмотрим следующую задачу. Имеется СМО с **n** каналами обслуживания. Очередь отсутствует. На вход системы поступает поток требований (заявок). В случае наличия свободного канала заявка поступает в систему на обслуживание. Если в момент поступления заявки все каналы заняты, то она получает <u>отказ</u>, покидает систему не обслуженной и в дальнейшем обслуживании не участвует. Системы подобного рода называются *<u>СМО без ожидания</u>*.

Пусть <u>входной поток</u> заявок - простейший, с интенсивностью - **λ** требований в единицу времени

<u>Время обслуживания</u> заявки $\mathbf{T}_{\text{обс}}$ подчинено экспоненциальному закону распределения с параметром $\mu = \frac{1}{t_{обс}}$. Здесь $\mathbf{t}_{\text{обс}}$ – среднее время (математическое ожидание) обслуживания требования в канале обслуживания. В этом случае параметр **μ** можно рассматривать как <u>интенсивность потока обслуживания</u> занятого канала (если он бесперебойно обеспечивается заявками на обслуживание), являющегося также простейшим потоком.

По классификации Кендалла такую СМО можно обозначить как M/M/n с длиной очереди равной нулю.

Составим граф состояний для данной системы.



Здесь состояние $\mathbf{S}_{\mathbf{k}}$ (k=$\overline{0, n}$) соответствует числу заявок, находящихся в системе на обслуживании и равно числу занятых каналов.

Переход системы из состояния **S_{k-1}** в состояние **S_k** происходит при поступлении очередной заявки с интенсивностью **$\lambda_{k-1,k}$** = **λ** , а из состояния S_k в состояние S_{k-1} - при обслуживании любой из заявок в одном из k каналов системы, занятых обслуживанием заявок, с суммарной интенсивностью **$\mu_{k,k-1}$** = **μ*k** .

Так как оба потока - входной и обслуживания – простейшие, то система является марковской, а процессы, в ней протекающие, могут быть описаны как процессы размножения и гибели с помощью уравнений Колмогорова-Чепмена.

$$\begin{cases} \dfrac{\partial P_o(t)}{\partial t} = -\lambda * P_o(t) + \mu * P(t)_1 \\ \dfrac{\partial P_1(t)}{\partial t} = -(\lambda + \mu) * P_1(t) + \lambda * P_0(t) + 2\mu * P_2(t) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{\partial P_k(t)}{\partial t} = -(\lambda + k * \mu) * P(t)_k + \lambda P_{k-1}(t) + (k+1) * \mu * P_{k+1}(t) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{\partial P_n(t)}{\partial t} = -n * \mu * P_n(t) + \lambda P_{n-1}(t) \end{cases}$$

Задав начальные условия и решив данную систему дифференциальных уравнений можно найти все $P_k(t)$ - вероятности того, что в момент времени **t** в системе находится **k** требований (занято **k** каналов обслуживания).

Значения финальных вероятностей найдем, приравняв $\dfrac{\partial P_k(t)}{\partial t}$ = 0 и решив соответствующую систему алгебраических уравнений.

Используя найденные ранее выражения для финальных вероятностей – получим:

$$P_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda \cdot \lambda}{\mu \cdot 2\mu} + \ldots + \frac{\lambda \cdot \lambda \ldots \lambda}{\mu \cdot 2\mu \cdot \ldots \cdot n\mu}\right)^{-1} = \left(\sum_{k=0}^{n} \frac{\lambda^k}{k! \cdot \mu^k}\right)^{-1}$$

Обозначим $\rho = \dfrac{\lambda}{\mu}$ - ***приведённая плотность потока заявок***. Её можно рассматривать как среднее число требований во входном потоке, приходящихся на среднее время обслуживания одного требования. Тогда

$$P_0 = (1 + \frac{\rho}{1} + \frac{\rho^2}{2!} + .. + \frac{\rho}{n!})^{-1} = (\sum_{k=0}^{n} \frac{\rho^k}{k!})^{-1}$$

Вероятности состояний, соответственно равны

$$P_k = (\frac{\lambda}{\mu})^k \cdot \frac{1}{k!} \cdot P_0 = \frac{\rho^k}{k!} \cdot P_0$$

Дифференциальные уравнения, описывающие поведение системы без ожидания называются **уравнениями Эрланга**, а выражения для финальных вероятностей - **формулами Эрланга**- в честь основателя ТМО - датского учёного Эрланга, чьи исследования телефонных линий в начале 20-го века, заложили основы современной Теории Массового Обслуживания.

Зная значения финальных вероятностей можно найти основные характеристики работы СМО в стационарном режиме.

1) Вероятность того, что пришедшее требование получит отказ

$$P_{отк} = P_n = \frac{\rho^n}{n!} \cdot P_0$$

2) Вероятность того, что пришедшее требование будет обслужено

$\boldsymbol{P_{обс}=1-P_{отк}=1-P_n}$

3) Абсолютная (фактическая) пропускная способность системы (среднее число требований, обслуженных за единицу времени)

$$A = \lambda \cdot P_{obs} = \lambda \cdot (1 - \frac{\rho^n}{n\ !} \cdot P_0)$$

4) <u>Максимальная пропускная способность системы</u> (максимальное среднее число требований, которое может обслужить система за единицу времени, если все каналы будут постоянно заняты)

$\boldsymbol{A_{max}=n*\mu}$.

5) ***<u>Среднее число занятых каналов</u>***

$$\mathrm{m}_{кан} = a_n = \sum_{k=0}^{n} k \cdot P_k \cdot \frac{\rho^k}{k!} \cdot P_0 = \rho \cdot P\,обс$$.

Так как каждый канал в единицу времени из **A** заявок, обнаруженных системой, обслуживает ровно **μ**, то <u>среднее число занятых каналов</u> может быть найдено как

$$m_{кан} = a_n = \frac{A}{\mu} = \frac{\lambda \cdot (1 - \frac{\rho^n}{n!} \cdot P_0)}{\mu} = \rho \cdot (1 - \frac{\rho^n}{n!} \cdot P_0) = \rho \cdot Pооб$$

6) ***<u>Интенсивность потока отказов</u>***

$$\lambda отк = \lambda \cdot P_{отк} = \lambda \cdot \frac{\rho^n}{n!} \cdot P_0$$

7) ***<u>Интенсивность потока обслуженных заявок</u>***

$\boldsymbol{\mu_{обс} = \mu * m_{кан} = \mu * \rho * P_{обс} = \lambda * P_{обс}}$

8) Как мы знаем, для рассматриваемой СМО, ***<u>среднее время нахождения заявки в системе</u>*** равно среднему времени нахождения заявки в канале обслуживания (очередь отсутствует). $\mathbf{t_{сист}} = \dfrac{1}{\mu}$.

Аналогично его значение можно получить и по формуле Литтла: $\mathbf{t_{сист}}$ = $\mathbf{m_{сист}} * \dfrac{1}{\lambda сист} = \mathbf{m_{сист}} * \dfrac{1}{\lambda \cdot P_{обс}} = \dfrac{1}{\mu}$.

Лекция №8.

Модели марковских СМО. СМО с ожиданием. СМО с неограниченной очередью

Рассмотрим теперь другую задачу. Имеется СМО с **n** каналами обслуживания. На вход системы поступает поток требований (заявок). В случае наличия свободного канала заявка поступает на обслуживание. Если в момент поступления заявки все каналы заняты, то заявка становится в очередь, ожидая пока освободится какой-либо канал обслуживания.

Такие системы называются системами с ожиданием. При этом время ожидания в очереди может быть как неограниченным («чистые системы с ожиданием»), так и ограниченным. Кроме того, длина очереди может быть как неограниченная, так и ограниченная. В случае заполнения всей очереди, пришедшие требования получают отказ и не обслуживаются.

Рассмотрим СМО с неограниченной очередью **(m=∞)** и неограниченным временем ожидания в очереди. Поток требований на входе системы - простейший с интенсивностью **λ**. Продолжительность обслуживания требования подчинена экспоненциальному закону распределения со средним временем обслуживания $m_t = \frac{1}{\mu}$. Тогда систему можно рассматривать как марковскую вида M/M/n. Изменение вероятностей состояния во времени может быть записано с помощью уравнений размножения и гибели.

$$\begin{cases} \dfrac{dP_0(t)}{dt} = -\lambda \cdot P_0(t) + \mu \cdot P_1(t) \\ \dfrac{dP_k(t)}{dt} = -(\lambda + k\mu) \cdot P_k(t) + \lambda \cdot P_{k-1}(t) + (k+1)\mu \cdot P_{k+1}(t), \quad 1 \le k < n \\ \dfrac{dP_k(t)}{dt} = -(\lambda + n\mu) \cdot P_k(t) + \lambda \cdot P_{k-1}(t) + n \cdot \mu \cdot P_{k+1}(t), \quad k \ge n \end{cases}$$

В этой системе **при** **k<n** интенсивность процесса «размножения» (переход от состояния $\mathbf{S_{k-1}}$ к состоянию $\mathbf{S_k}$) всегда равна **λ**, а интенсивность процесса «гибели» (переход от состояния $\mathbf{S_{k+1}}$ к состоянию $\mathbf{S_k}$) равна **(k+1)*μ**.

При k≥n интенсивность процесса «размножения» остается равной **λ**, а интенсивность процесса «гибели» всегда равна **n*μ** (так как все каналы обслуживания заняты) и не зависит от числа требований в системе.

Приравнивая все $\frac{dP_k(t)}{dt}=0$, получим для стационарного режима выражения для финальных вероятностей.

$$P_k = \frac{\rho^k}{k!}\cdot P_0, \quad 1\le k\le n$$

$$P_k = (\frac{\rho}{n})^{k-n}\cdot P_n = \frac{\rho^k}{n!\,n^{k-n}}\cdot P_0, \quad k\ge n$$

Вероятность $\mathbf{P_0}$ ищется из нормирующего условия и равна

$$P_0 = (1+\sum_{k=1}^{n-1}\frac{\rho^k}{k!}+\frac{\rho^n}{n!}\sum_{k=n}^{\infty}\frac{\rho^{k-n}}{n^{k-n}})^{-1}$$

Условием существования стационарного режима и, соответственно, ненулевых P_k является сходимость бесконечного ряда в выражении для P_0. Для этого необходимо, чтобы интенсивность процесса "размножения", начиная с **k=n**, была бы меньше, чем интенсивность процесса гибели, т.е. $\frac{\lambda}{n\mu}=\frac{\rho}{n}<1$,

или, иначе, **(ρ < n)**.

Если это условие не выполняется, то режима статического равновесия в СМО не существует, что приводит к неограниченному увеличению очереди. Это условие необходимо обязательно учитывать при проектировании реальных СМО.

Если условие (**ρ/n<1)** выполняется, то, используя выражение для суммы членов бесконечной убывающей геометрической прогрессии, получим для P_0

$$\sum_{k=n}^{\infty}\left(\frac{\rho}{n}\right)^{k-n} = \frac{1}{1-\frac{\rho}{n}}$$

Отсюда

$$P_0 = [1 + \sum_{k=1}^{n-1}\frac{\rho}{k!} + \frac{\rho^n}{n!} * \frac{1}{1-\frac{\rho}{n}}]^{-1}$$

Найдем теперь основные характеристики СМО с неограниченной очередью в стационарном режиме:

1) Так как очередь не ограничена и все требования, поступающие в систему будут обслужены, то ***вероятность обслуживания***

$\boldsymbol{P_{обс} = 1}$, а вероятность отказа в обслуживании $P_{отк} = \mathbf{0}$.

2) **П$_n$** – ***вероятность того, что все каналы*** обслуживания будут ***заняты***

$$\Pi_n = \sum_{k=n}^{\infty} P_k = \sum_{k=n}^{\infty}\left(\frac{\rho}{n}\right)^{k-n} * P_n = \frac{1}{1-{}^{\rho}/_{n}} * P_n$$

или через P_0 $\quad \Pi_n = \frac{1}{1-{}^{\rho}/_{n}} * \frac{\rho^n}{n!} * P_0$.

3) ***Среднее число занятых каналов***

$$m_{кан} = a_n = \sum_{k=0}^{n-1} k * P_k + n * \Pi_n = \left(\sum_{k=0}^{n-1} k * \frac{\rho^k}{k!} + \frac{n}{1-{}^{\rho}/_{n}} * \frac{\rho^n}{n!}\right) * P_0$$

Т.к. все требования, поступающие в систему, обслуживаются ($P_{обс}$ =1), то абсолютная пропускная способность системы (число требований обслуживаемых в системе в единицу времени) **A=λ**. Отсюда **среднее число занятых каналов** можно найти как

$$\mathbf{m_{кан}} = \boldsymbol{a_n = \frac{A}{\mu} = \frac{\lambda}{\mu} = \rho} \; .$$

Т.е. <u>в стационарном режиме</u>, при условии, что **(ρ<n)** <u>среднее число занятых каналов</u> **$m_{кан}$** <u>всегда равно **ρ**</u> и не зависит от числа каналов.

4) ***<u>Средняя длина очереди</u>***.

$$m_L = \sum_{k=n}^{\infty}(k-n)*P_k = P_n*\sum_{l=0}^{\infty} l*\left(\frac{\rho}{n}\right)^l$$

При $\frac{\rho}{n} < 1$ т.е.(**ρ<n**) ряд сходится и равен

$$\sum_{l=0}^{\infty} l*\left(\frac{\rho}{n}\right)^l = \frac{\frac{\rho}{n}}{\left(1-\frac{\rho}{n}\right)^2}$$

Отсюда <u>средняя длина очереди</u> будет равна

$$m_L = P_n*\frac{\frac{\rho}{n}}{\left(1-\frac{\rho}{n}\right)^2} = P_0*\frac{\rho^n}{n!}*\frac{\frac{\rho}{n}}{\left(1-\frac{\rho}{n}\right)^2}$$

5) ***<u>Среднее число требований, находящихся в системе</u>***

$$m_k == \sum_{k=0}^{\infty} k*P_k = m_L + \rho = m_{кан} + m_L$$

6) Используя формулу Литтла можно найти ***<u>среднее время нахождения требования в очереди и в системе</u>***.

$$t_{оч} = \frac{1}{\lambda_{сист}}*m_L = \frac{1}{\lambda}*P_n*\frac{\frac{\rho}{n}}{\left(1-\frac{\rho}{n}\right)^2} = \frac{1}{\rho\mu}*m_L$$

$$t_{сист} = \frac{1}{\lambda_{сист}}*m_{сист} = \frac{1}{\lambda}(m_L + a_n) = \frac{1}{\lambda}(m_L + \rho)$$

Отсюда $t_{сист} = t_{оч} + t_{кан} = \frac{1}{\rho\mu}*m_L + \frac{1}{\mu}$

СМО с ожиданием. СМО с ограниченной длиной очереди.

Рассмотрим **n**-канальную СМО с очередью, максимальная длина которой ограничена и равна **m**. Время ожидания в очереди – не ограниченно.

Поток требований на входе в систему простейший с интенсивностью $\boldsymbol{\lambda}$. Поток обслуживания также простейший с интенсивностью $\boldsymbol{\mu}$. В этом случае система будет марковской вида **M/M/n/m** и изменения вероятностей состояния могут быть описано с помощью уравнений размножения и гибели вида

$$\frac{dP_o(t)}{dt} = -\lambda * P_o(t) + \mu * P_1(t)$$

$$\frac{dP_k(t)}{dt} = -(\lambda + k\mu) * P_k(t) + \lambda P_{k-1}(t) + (k+1)\mu * P_{k+1}(t) \quad \underline{1 \le k < n}$$

$$\frac{dP_k(t)}{dt} = -(\lambda + n\mu) * P_k(t) + \lambda P_{k-1}(t) + n * \mu * P_{k+1}(t) \quad \underline{n \le k < n+m}$$

$$\frac{dP_{n+m}(t)}{dt} = -n * \mu * P_{n+m}(t) + \lambda * P_{n+m-1}(t)$$

Т.к. число состояний системы конечное, то в ней существует стационарный режим, и выражения для финальных вероятностей могут быть записаны в следующем виде:

$$\mathbf{P_k = \frac{\rho^n}{k!} * P_0, \quad 1 \le k \le n}$$

$$\mathbf{P_k = \left(\frac{\rho}{k}\right)^{k-n} * P_n = \frac{\rho^n}{n!} * \left(\frac{\rho}{k}\right)^{k-n} * P_0\,, \quad n \le k \le n+m,}$$

где $\mathbf{P_0}$, найденная из нормирующего условия $\sum_{k=0}^{n+m} \mathbf{P_k = 1}$,будет равна

$$P_0 = [1 + \sum_{k=1}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n}{k!} \sum_{k=n}^{n+m} \left(\frac{\rho}{n}\right)^{k-n}]^{-1}$$

Учитывая, что вторая сумма, как сумма членов геометрической прогрессии равна $\sum_{k=n}^{n+m}\left(\frac{\rho}{n}\right)^{k-n} = \frac{1-(^{\rho}/_{n})^{m+1}}{1-^{\rho}/_{n}}$, получим окончательно

$$P_0 = [1 + \sum_{k=1}^{n-1}\frac{\rho^k}{k!} + \frac{\rho^n}{k!} * \frac{1-(^{\rho}/_{n})^{m+1}}{1-^{\rho}/_{n}}]^{-1}$$

<u>Основные характеристики СМО в стационарном режиме будут:</u>

1) ***<u>Вероятность</u>*** того, ***<u>что</u>*** пришедшее требование ***<u>получит отказ</u>***

$$P_{отк} = P_{n+m} = (\frac{\rho}{n})^m * P_n = \frac{\rho^n}{n!}(\frac{\rho}{n})^m * P_0$$

2) ***<u>Вероятность того, что пришедшее требование будет обслужено</u>***.

$$P_{обс} = 1 - P_{отк} = 1 - \frac{\rho^n}{n!}(\frac{\rho}{n})^m * P_0$$

3) ***<u>Абсолютная пропускная способность системы</u>*** (среднее число требований, обслуженной за единицу времени)

$$A = \lambda * P_{обс} = \lambda * (1 - \frac{\rho^n}{n!}\left(\frac{\rho}{n}\right)^m * P_0)$$

4) ***<u>Вероятность того, что все каналы обслуживания заняты</u>***.

$$\pi_n = \sum_{k=n}^{n+m} P_k = \sum_{k=n}^{n+m}\left(\frac{\rho}{n}\right)^{k-n} = P_0 * \frac{1-(^{\rho}/_{n})^{m+1}}{1-^{\rho}/_{n}} * \left(\frac{\rho}{n!}\right)^n$$

5) ***<u>Среднее число занятых каналов</u>***.

$$m_{кан} = a_n = \sum_{k=1}^{n} k * P_k + n * \sum_{k=n+1}^{n+m} P_k$$

Или, иначе: $m_{кан} = \rho * P_{обс}$

6) ***<u>Средняя длина очереди</u>***.

$$\mathbf{m_L} = \sum_{k=n+1}^{n+m}(k-n) * \mathbf{P_k} = \frac{\rho^n}{n!} * \mathbf{P_0} * \frac{\rho/_n}{(1-\rho/_n)^2} * \left[1-(m+1) * \left(\frac{\rho}{n}\right)^m + m * \left(\frac{\rho}{n}\right)^{m+1}\right]$$

7) ***<u>Среднее число требований в системе</u>*** можно найти как

$$\mathbf{m_{сист}} = \mathbf{a_n} + \mathbf{m_L} = \mathbf{m_L} + \rho * \mathbf{P_{обс}}\ .$$

8) ***<u>Среднее время нахождения требования в очереди</u>***

$$\mathbf{t_{оч}} = \frac{1}{\lambda_{сист}} * \mathbf{m_L} = \frac{1}{\lambda * P_{обс}} * \mathbf{m_L}$$

9) ***<u>Общее время нахождения требования в системе</u>*** будет

$$\mathbf{t_{сист}} = \mathbf{t_{оч}} + \mathbf{t_{кан}} = \frac{1}{\lambda_{сист}}(\mathbf{m_L} + \mathbf{a_n}) = \frac{1}{\lambda * \mathbf{P_{обс}}} * [\mathbf{m_L} + \rho * \mathbf{P_{обс}}] = \mathbf{t_{оч}} + \frac{1}{\mu}$$

Полученные выражения носят общий характер. Положив в них **m=0** получим значения $P_к$ и характеристики СМО для систем без ожидания. Положив **m=∞,** получим выражения для СМО с неограниченным числом мест в очереди.

<u>Пример.</u> Ателье обслуживает жителей 2-х микрорайонов. Интенсивности заявок с обоих микрорайонов равна $\lambda_1 = \lambda_2 = 2$ чел/час. Среднее время выполнения заявки составляет 20 мин. Число работников в ателье 2 человека. Определить основные характеристики работы ателье, если:

1) Очереди невозможны (m=0).

2) Очереди не ограничены (m=∞).

3) Как изменится эффективность работы, если каждый из работников будет обслуживать клиентов только одного района.

1) <u>Рассмотрим случай, когда</u> <u>очередь отсутствует</u> (**m=0**).

а) *Каждый работник обслуживает жителей любого района*. Тогда:

n=2; $\boldsymbol{\lambda = \lambda_1 + \lambda_2}$=4; $\boldsymbol{\mu = \frac{1}{1/3} = 3}$; $\boldsymbol{\rho = 4/3}$;

$P_0 = [1 + \sum_{k=1}^{n} \frac{\rho^k}{k!}]^{-1} = \frac{9}{29}$; $P_k = \frac{\rho^k}{k!}$;

$P_1 = \frac{\rho}{1} * \rho_0 = \frac{12}{29}$; $P_2 = \frac{\rho^2}{2!} * \rho_0 = \frac{8}{29}$;

$P_{отк} = P_2 = \frac{8}{29}$; $P_{обс} = 1 - P_{отк} = \frac{21}{29}$=0,724;

A= $\lambda * P_{обс} = 4 * \frac{21}{29}$=2,9; $a_n = \frac{A}{\mu} = 0{,}96$;

б) *Каждый работник обслуживает жителей только одного района*. Тогда:

n=1; $\boldsymbol{\lambda = \lambda_1 = \lambda_2}$=2; **μ=3**; $\boldsymbol{\rho = 2/3}$;

$P_0 = \frac{3}{5}$; $P_1 = \frac{\rho}{1} * \rho_0 = \frac{2}{5}$;

$P_{отк} = P_1 = 2/5$, $P_{обс} = 1 - P_{отк} = \frac{3}{5} = 0{,}6$;

A= $\lambda * P_{обс}$=2*3/5=1,2; $a_n = \frac{A}{\mu} = 0{,}4$;

Как видим, во втором случае снижается вероятность обслуживания каждого клиента (с 0,724 до 0,6) и общее число обслуживаемых клиентов A=2*1,2=2,4 (меньше чем 2,9). Загрузка мастеров так же падает

$a_n = 2 * 0{,}4 = 0{,}8$ (меньше чем 0,96).

2) <u>Рассмотрим случай, когда очередь не ограничена (**m=∞**).</u>

а) *Каждый работник обслуживает жителей любого района*. Тогда:

n=2; $\boldsymbol{\lambda = \lambda_1 + \lambda_2}$=4; **μ=3**; $\boldsymbol{\rho = 4/3}$;

$\boldsymbol{\frac{\rho}{n} = 2/3 < 1}$ (стационарный режим существует);

$P_0 = \left[1 + \sum_{k=1}^{n} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} * \frac{1}{1 - \frac{\rho}{n}}\right]^{-1} = \frac{3}{15}$; $P_{обс} = 1$;

$P_k = \frac{\rho^k}{k!} * P_0 \; 1 \le k \le n$; $\qquad P_1 = \frac{4}{15}$; $\qquad P_2 = \frac{8}{45}$;

$P_k = \left(\frac{\rho}{n}\right)^{k-n} * P_n k \ge n$: $\; P_3 = \frac{2}{3} * \frac{8}{45}$; $P_4 = \left(\frac{2}{3}\right)^2 * \frac{8}{45} \ldots$

$\pi_n = \sum_{k=n}^{n+m} P_k = P_n * \frac{1}{1-\frac{\rho}{n}} = \frac{8}{15}$;

$A = \lambda * P_{обс} = \lambda = 4$; $\qquad a_n = \frac{A}{\mu} = \rho = \frac{4}{3}$;

$m_L = P_n * \frac{\rho/n}{(1-\rho/n)^2} = 1{,}07$; $\quad t_{оч} = \frac{1}{\lambda} * m_L = 0{,}28 \text{часа} = 17 \text{минут}$.

б) *Каждый работник обслуживает жителей только одного района.* Тогда:

$n = 1$. $\qquad \lambda = 2$; $\qquad \mu = 3$; $\qquad \rho = \frac{2}{3}$

$P_0 = (1 + \rho)^{-1} = \frac{1}{3}$ $\qquad P_k = P_n \cdot \left(\frac{\rho}{n}\right)^k \; k>n$

$P_1 = \rho * P_0 = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$, $P_2 = \rho * P_1 = \frac{2}{3} \cdot \frac{2}{9} = \frac{4}{27}$, $\quad P_3 = \left(\frac{2}{3}\right)^2 \cdot \frac{2}{9} = \frac{8}{81}$, ……

$\Pi_n = P_n \frac{1}{(1-3/4)} = \frac{2}{9} \frac{1}{\left(1-\frac{2}{3}\right)} = \frac{2}{3}$, $\quad a_n = \rho = \frac{2}{3}$

$m_L = \frac{4}{3} \approx 1.33$, $\quad t_{оч} = \frac{1}{\lambda} \cdot m_L = \frac{1}{2} \cdot 1.33 = 0.665 \text{ часа} \approx 40 \text{ мин.}$

В случае неограниченной очереди (все жители обслужены) при том же среднем числе загруженных каналов $\frac{4}{3} = 2 \cdot \frac{2}{3}$ увеличивается средняя длина очереди (1.33 в каждом ателье против 1.07 – общая очередь в ателье с 2-мя мастерами). Среднее время ожидания в очереди при этом возросло более чем в 2 раза с 17 мин. до 40 мин.

Как видно, разделение СМО на несколько систем, работающих независимо, с разделением потока заявок приводит к снижению эффективности работы в целом.

Лекция № 9.

СМО с ограничением времени ожидания в очереди.

Пусть система имеет **n** каналов обслуживания. Входящий поток требований – простейший с интенсивностью **λ**. Время обслуживания распределено по экспоненциальному закону с параметром **μ**. Если в момент поступления требования все каналы заняты, то оно встает в очередь. Однако время пребывания требования в очереди – ограничено. По окончанию этого времени, если требование не поступило на обслуживание, то оно покидает систему не обслуженным.

Следует отметить, что ограничение на время ожидания в очереди носит принципиальный характер, т.к. при вычислении вероятностей состояний СМО необходимо знание не только текущего состояния (числа требований в системе), но и того, как давно пришли требования, ожидающие обслуживания. Таким образом, процесс функционирования системы характеризуется наличием последействия и, в общем случае, перестает быть марковским. При этом *максимальное время ожидания может быть как детерминированным, так и случайным*.

Рассмотрим случай, когда максимальное время ожидания T = const. В этом случае вероятность перехода системы из одного состояния в другое зависит не только от того, сколько требований находится в системе на обслуживании и в очереди (т.е. от состояния системы), но и от того, как давно требование, стоящее в очереди, находится в системе. От этого зависит – будет ли требование оставаться в системе или покидает ее, как только время ожидания превысит допустимое. Т.е. *налицо последействие* и последовательность состояний системы (процесс изменения состояний) – **не марковский.**

Для описания процесса функционирования такой СМО нельзя, напрямую, использовать уравнения размножения и гибели. Однако, описание СМО с помощью марковской модели возможно, если использовать расширение понятия состояния. Для того, чтобы прогнозировать распределения состояний в будущем, необходимо знать, как давно пришли в систему требования, которые в настоящее время находятся в очереди. Это можно сделать, включив в число обобщенных координат, описывающих состояние СМО, давность прихода каждого ожидающего требования, или, что то же самое, время, которое осталось у него до окончания срока ожидания. Расширение понятия состояния приводит к увеличению мерности задачи при учете состояний системы. (см. конспект Бомас.)

Для исследования подобных СМО как немарковских (как, впрочем, и любых других, в том числе и марковских) широко используется имитационное моделирование. При этом оценки параметров исследуемой СМО ищутся на основе статистической обработки результатов моделирований, полученных по множеству прогонов (реализаций).

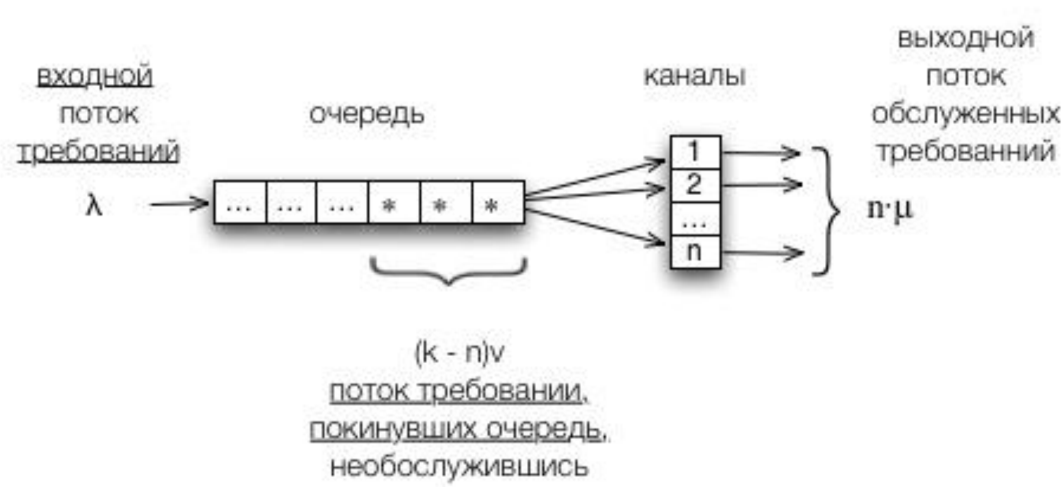
Пусть ***Максимальная длительность ожидания ограничена случайной величиной.*** В этом случае все зависит от закона распределения ограничения максимального времени ожидания T, т.к. именно это ограничение вносит в систему последействие. Поэтому вернуть процессу функционирования СМО марковость крайне просто. Достаточно принять для описания случайной величины T экспоненциальное распределение. Однако при этом нельзя забывать, что такая операция возможна лишь в том случае, когда реальное распределение или действительно экспоненциальное, или близко к нему. Если это не так, то сформированная математическая модель будет неадекватной реальной СМО.

Ограничимся случаем, когда максимальная длительность ожидания требования в очереди $\boldsymbol{T}$ имеет экспоненциальное распределение: $F(\tau) = 1 - e^{-\nu\tau}\tau \geq 0$ (функция распределения), $f(\tau) = \boldsymbol{\nu} e^{-\nu\tau}$ (функция плотности). Здесь **ν** – параметр закона распределения. Среднее значение τ_{max} будет $m_\tau = \frac{1}{\nu}$.

Такие процессы функционирования СМО можно представить следующим образом (см. рис.).

На вход системы поступает поток требований (простейший) с интенсивностью $\boldsymbol{\lambda}$. *Если* не все каналы заняты **(k < n)**, то пришедшие требования сразу же начинают обслуживаться. Обслуженные требования покидают систему с интенсивностью **k · μ**.

Если все каналы системы заняты, **(k > n)**, то обслуженные требования покидают систему с интенсивностью **n · μ**. Кроме того, возникает еще один поток требований – покидающих систему не дождавшись обслуживания. Интенсивность этого потока равна **(k – n) · ν** (см. рис.).



Т.к. показательное распределение максимального времени ожидания также обладает свойством «отсутствия памяти», т.е. распределение оставшегося времени ожидания не зависит от того, сколько времени требование уже прождало в очереди, то последовательность состояний системы оказывается «без последействия» и является **«Марковской»**. Это позволяет описать ее с помощью уравнений «размножения и гибели». В частности, для неограниченной очереди **(m = ∞)** они имеют вид:

$$\frac{dP_0}{dt} = -\lambda \cdot P_0 + \mu \cdot P_1$$

$$\frac{dP_K}{dt} = -(\lambda + k \cdot \mu) \cdot P_k + \lambda \cdot P_{k-1} + (k+1) \cdot \mu \cdot P_{k+1} \quad 1 \le k < n$$

..

$$\frac{dP_K}{dt} = -[\lambda + n\mu + (k-n) * v] * P_k + \lambda P_{k-1} +$$

$$[n\mu + (k+1-n) * v] * P_{k+1} \; , \qquad\qquad k \ge n \, .$$

Если в системе существует стационарный режим, то значения финальных вероятностей соответственно равны:

$$P_{k+1} = \frac{\lambda_{k,k+1}}{\mu_{k+1,k}} \; , \quad P_k = \frac{\rho^k}{k!} \cdot P_0 , \;\; k \le n \, ,$$

где $P_0 = \left[1 + \sum_{k=1}^{n} \frac{\rho^k}{k\,!} + \frac{\rho^n}{n\,!} * \sum_{k=n+1}^{\infty} \frac{\lambda^{(k-n)}}{\prod_{l=1}^{k-n} (l\nu + n\mu)} \right]^{-1}$.

Зная значения финальных вероятностей можно найти **основные характеристики функционирования СМО в стационарном режиме** ($m_{кан}$; $m_{оч}$; $m_{отк}$ и другие).

При определении характеристик функционирования СМО необходимо иметь в виду, что не все требования, поступившие в систему, будут обслужены.

1) Вероятность того, что требование, подошедшее к системе, получит отказ и не попадет в систему

$$\mathbf{P_{отк} = P_{n+m}}$$

2) Вероятность того, что подошедшее требование поступит в систему

$$\mathbf{P_{сист} = 1 - P_{отк} = 1 - P_{n+m}}\ .$$

Поток этих требований с интенсивностью

$\boldsymbol{\lambda_{сист} = \lambda * P_{сист} = \lambda*(1-P_{n+m})}$ разобьётся на два потока:

- обслуженных требований, с интенсивностью $\boldsymbol{\lambda_{обс} = m_{кан} * \mu}$;

- покинувших систему не обслуженными, из-за ограничения на время ожидания в очереди, с интенсивностью $\boldsymbol{\lambda_{пок} = m_{оч} * \nu}$.

Причем, **в стационарном режиме** для этих потоков будет выполняться условие

$$\boldsymbol{\lambda_{сист} = \lambda_{обс} + \lambda_{пок}}\ .$$

Тогда для любого требования, вошедшего в систему из входного потока, вероятности обслуживания $\boldsymbol{p_{обс}}$ или не обслуживания (покидания системы из очереди не обслуженными) $\boldsymbol{p_{пок}}$ *в стационарном режиме* будут пропорциональны интенсивностям соответствующих потоков:

$$P_{обс}^{сист} = \frac{\lambda_{обс}}{\lambda_{сист}} = \frac{m_{кан}\mu}{\lambda_{сист}}, \quad P_{пок}^{сист} = \frac{\lambda_{пок}}{\lambda_{сист}} = \frac{m_{оч}\nu}{\lambda_{сист}}.,$$

а для требования, подошедшего на входе к системе, соответствующие вероятности будут равны:

$$P_{обс}^{вх} = P_{сист}\frac{\lambda_{обс}}{\lambda_{сист}} = \frac{P_{сист}m_{кан}\mu}{\lambda P_{сист}} = \frac{m_{кан}\mu}{\lambda}, \quad P_{пок}^{вх} = P_{сист}\frac{\lambda_{пок}}{\lambda_{сист}} = \frac{P_{сист}m_{оч}\nu}{\lambda P_{сист}} = \frac{m_{оч}\nu}{\lambda}$$

Следует отметить, что в режиме статистического равновесия всегда должно выполняться условие равенства потоков требований

$$\boldsymbol{\lambda = \lambda_{отк} + \lambda_{сист} = \lambda_{отк} + \lambda_{обс} + \lambda_{пок}}.$$

<u>Средние временные характеристики для требований</u>, на основании формулы Литтла будут:

1. <u>Среднее время нахождения в канале обслуживания</u> $t_{кан} = \dfrac{m_{кан}}{\lambda_{обс}} = \dfrac{m_{кан}}{m_{кан}\mu} = \dfrac{1}{\mu}$
2. <u>Среднее время нахождения в очереди</u> $t_{оч} = \dfrac{m_{оч}}{\lambda_{сист}} = \dfrac{m_{оч}}{\lambda обс + \lambda пок} = \dfrac{m_{оч}}{m_{оч}\nu + m_{кан}\mu}$
3. <u>Среднее время нахождения в системе</u>

$$t_{сист} = t_{оч} + t_{кан} = \frac{m_{оч}}{m_{оч}\nu + m_{кан}\mu} + \frac{1}{\mu}$$

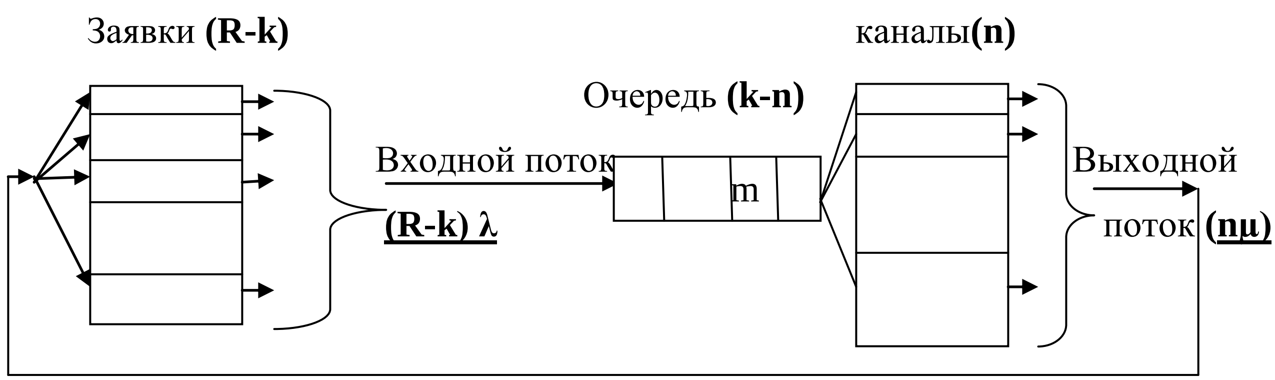
Замкнутые СМО

Наряду с задачами, когда интенсивность входного потока требований в систему **λ** неизменна, на практике часто встречаются задачи, когда число источников требований ограничено, и интенсивность входного потока зависит от того, сколько источников уже послало требования в систему на обслуживание. В качестве примера рассмотрим следующую задачу.

Система содержит **R** станков, которые обслуживают **n** рабочих. Каждый станок может отказать в случайный момент времени. Время безотказной работы каждого станка подчинено экспоненциальному закону распределения $F(t) = 1 - e^{-\lambda t}, t \geq 0$, где **t** – момент наступления отказа, **λ** – параметр, который можно рассматривать как интенсивность потока отказов одного станка, если считать, что после отказа он мгновенно переводится в рабочее состояние и снова работает до очередного отказа.

Если хотя бы один из рабочих свободен, то отказавший станок начинает обслуживаться (ремонтируется). Время обслуживания распределено также экспоненциально с параметром **μ**. Если все рабочие заняты, то станок становится в очередь на обслуживание (ремонт), длина которой не может быть больше чем **m = R - n** .

Схематически данную систему, как СМО можно представить в следующем виде:



Обозначим **k** – число заявок на обслуживание (число отказавших станков), **0 ≤ k ≤ R**.

Тогда интенсивность входного потока заявок зависит от числа отказавших станков и будет равна **(R-k)·λ.** С учётом этого уравнения Колмогорова-Чепмена, которыми можно описать данную СМО, примут вид (**при условии, что R>n**):

$$\frac{dP_0}{dt} = -R\lambda P_0 + \mu P_1$$

$$\frac{dP_k}{dt} = -\left[\left(R-k\right)\lambda + k\mu\right]P_k + \left(R-\left(k-1\right)\right)\lambda P_{k-1} + \left(k+1\right)\mu P_{k+1},\ 1 \le k < n$$

$$\frac{dP_k}{dt} = -\left[\left(R-k\right)\lambda + n\mu\right]P_k + \left(R-\left(k-1\right)\right)\lambda P_{k-1} + n\mu P_{k+1},\ n \le k < R$$

$$\frac{dP_R}{dt} = -n\mu \cdot P_R + 1 \cdot \lambda \cdot P_{R-1}\ ,\ n \le k < R.$$

Так как общее число состояний системы конечно, то при $\mathbf{t} \to \infty$ в ней существует режим статистического равновесия. Приравняв правые части уравнений нулю можно найти значения финальных вероятностей:

$$P_k = \frac{R\left(R-1\right)\cdot\ldots\cdot\left[R-\left(k-1\right)\right]\lambda^k}{1\cdot 2\cdot\ldots\cdot k\cdot\mu^n}P_0^0 = \frac{\rho^k}{k!}\frac{R!}{\left(R-k\right)!}P_0,\quad 1 \le k < n$$

$$P_k = \frac{R\left(R-1\right)\cdot\ldots\cdot\left[R-\left(k-1\right)\right]\lambda^k}{1\cdot 2\cdot\ldots\cdot k\cdot\mu^n}\cdot\frac{\left(R-n\right)\cdot\ldots\cdot\left(R-\left(k-1\right)\right)\lambda^{k-n}}{n\cdot\ldots\cdot n\cdot\mu^{k-n}}P_0^0 = \frac{\rho^k}{k!,n^{k-n}}\frac{R!}{\left(R-k\right)!}P_0,\ n \le k \le R$$

$$P_0 = \left[1 + \sum_{k=1}^{n-1}\frac{\rho^k}{k!}\frac{R!}{\left(R-k\right)!} + \sum_{k=n}^{R}\frac{\rho^k}{n!n^{k-n}}\frac{R!}{\left(R-k\right)!}\right]^{-1}.$$

Зная значения финальных вероятностей можно найти основные характеристики работы СМО в стационарном режиме, например:

1) Среднее число отказавших станков

$\mathbf{M_{отк}} = \sum_{k=0}^{R} \mathbf{k} * \mathbf{P_k}$

2) Среднее число исправных станков

$\mathbf{M_{испр} = R - M_{отк}}$

3) Среднее число станков, находящихся в ремонте

$\mathbf{M_{рем} = \sum_{k=1}^{n} k * P_k + n \sum_{k=n+1}^{R} P_k}$

4) Среднее число неисправных станков, ожидающих ремонта (в очереди)

$\mathbf{M_{ож.рем} = \sum_{k=n+1}^{R} (k - n) * P_k}$

Все рассматриваемые выше СМО мы анализировали в предположении, что все потоки событий, происходящих в системе – простейшие. Это позволяло описывать их достаточно просто с помощи уравнений Колмогорова - Чепмена и, в частности, уравнений размножения и гибели. Кроме того – получать достаточно простые выражения для финальных вероятностей состояний системы и её основных характеристик в режиме статистического равновесия.

Лекция № 10.

Оптимизация параметров СМО

При организации СМО важно выбрать ее параметры так, чтобы наилучшим образом решать стоящие перед ней задачи. При этом качество их решений определяется, как правило, с помощью векторного критерия эффективности, компонентами которого являются частные показатели эффективности СМО.

Решение задачи выбора рациональных параметров СМО как векторной – затруднительно. Поэтому на практике достаточно часто используют сведение частных показателей эффективности в один – обобщенный с помощью различных процедур свертки и далее задачу рассматривают как монокритериальную.

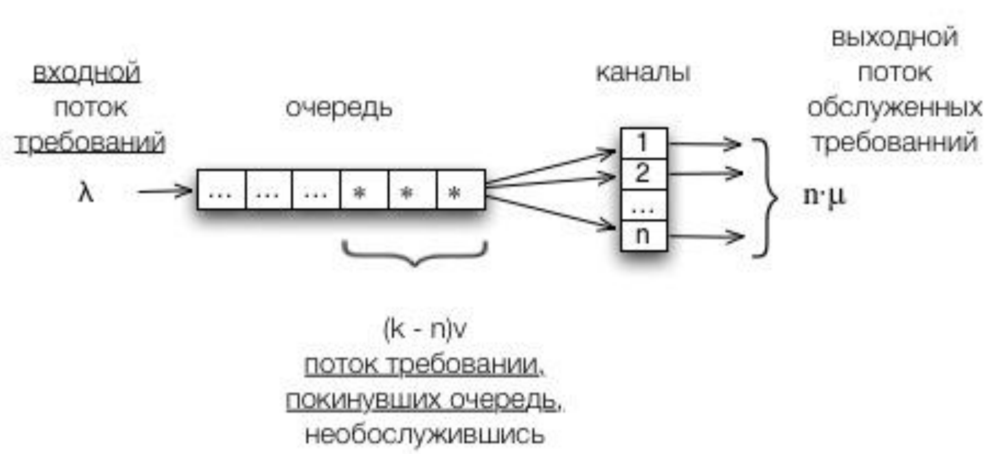
В качестве такого интегрального (обобщенного) показателя в широком классе задач можно использовать величину **прибыли**, получаемой от функционирования СМО, которая определяется как разница между доходами от обслуживания заявок (например, клиентов в парикмахерской), с одной стороны, и расходами на создание (модернизацию) системы (создание каналов обслуживания, мест в очереди, аренда помещений, зарплата работников, увеличение интенсивности обслуживания и т.п.),– расходами (штрафами, бонусами), которые вынуждена нести система, чтобы «не отпугнуть» клиентов за отказ в обслуживании, за незавершенное обслуживание) и т.д., с другой стороны.

Оптимизируемыми параметрами при этом могут быть: число каналов обслуживания **n**, максимальная длина очереди **m**, интенсивность обслуживания $\boldsymbol{\mu}$, стоимостные показатели и другие параметры СМО, которые варьируются в рамках заданных для них ограничений.

Однако стоимостной критерий не является универсальным. В ряде задач большее значение играет факт выполнение СМО поставленной перед ней задачей. Например, если в качестве СМО рассматривается система ПВО. Тогда в качестве показателя эффективности такой системы можно рассматривать математическое ожидание числа обслуженных (пораженных) самолетов противника, вероятность проникновения самолета через систему ПВО (вероятность, что заявка не будет обслужена) и др.

Рассмотрим, в качестве примера, некоторую ***СМО* с ограничением времени ожидания требования в очереди** (см. рис.). Если все потоки в системе будут распределены по экспоненциальному закону, то такая СМО является марковской и ее процесс функционирования может быть описан с помощью уравнений Колмогорова-Чепмена (см. лекции – СМО с ограничением времени нахождения заявки в очереди).

Если условие марковости в системе не выполняются, то для исследования подобных СМО на практике широко используется имитационное моделирование. Следует отметить, что в этом случае в алгоритмы функционирования собственно системы можно включать различные варианты управления в зависимости от состояние внешней среды. При этом оценки параметров исследуемой СМО ищутся на основе статистической обработки результатов моделирований, полученных по множеству прогонов (реализаций).



Если число мест в очереди ограничено, то в системе существует стационарный режим.

Примером такой СМО может являться ***парикмахерская***. В качестве критерия эффективности будем использовать величину **прибыли** (в рублях), получаемой от функционирования данной СМО в стационарном режиме, которая определяется как разность между доходами от обслуживания клиентов (заявок) и расходами на содержание системы за некоторый заданный период времени **T** (например, один месяц). Оценим эффективность работы подобной системы в стационарном режиме.

Для проведения расчетов примем:

n – число каналов обслуживания;

m – максимальное число мест в очереди;

λ - интенсивность поступления заявок в единицу времени (в час);

μ - интенсивность обслуживания (среднее число заявок, обслуживаемых в канале обслуживания за один час);

ν- интенсивность покидания очереди (для заявки в очереди) ввиду ограничения максимального времени нахождения заявки в очереди (в час).

$S_д$ – средний доход от обслуживания одной заявки (руб.);

$S_{отк}$ – расходы (бонусы), которые представляет система клиентам (заявкам), не вошедшим в систему, если все места в очереди заняты (руб.);

$S_{пок}$ – расходы (бонусы), которые представляет система клиентам, вынужденным покинуть систему не обслуженными в связи с большой очередью и ограничением на время нахождения в ней (руб.);

$S_{ож}$ – расходы (чай, кофе,…), которые представляет система клиентам во время ожидания ими начала обслуживания (руб.);

$S_{кан}$ – расходы, которые несет система, для создания одного канала обслуживания, отнесенные к заданному периоду времени **T** (руб.);

$\mathbf{S_{оч}}$ – расходы, которые несет система для создания одного места в очереди, отнесенные к заданному периоду времени **T** (руб.).

Запишем для нашей СМО уравнения размножения и гибели, найдем значения всех P_k – х в стационарном режиме (см. лекции - СМО с ограничением времени ожидания требования в очереди).

В качестве основных характеристик СМО примем:

$p_{отк} = p_{n+m}$ - вероятность получения отказа.

$\lambda_{отк} = \lambda * p_{отк}$ - поток отказов в обслуживании.

$\lambda_{сист} = \lambda * (1 - p_{отк})$ - поток требований, вошедших в систему.

$\nu_{пок} = \nu * m_{оч}$ - поток требований, покидающих очередь по ограничению времени.

$\mu_{обс} = m_{кан} * \mu$ - поток обслуженных заявок..

$P_{обс}$ - вероятность того, что требование из входного потока будет обслужено.

$P_{пок}$ - вероятность, что требование из входного потока покинет очередь не обслуженным.

$t_{ож.оч.}$ - среднее время ожидания требования в очереди.

После проведенных расчетов необходимо убедиться, что в стационарном режиме работы системы выполняется условие:

$\lambda = \lambda отк + \lambda сист$, $\lambda сист = \mu обс + \nu пок$.

Примечание: При вычислении доходов и расходов в системе за заданный период времени **T** (например, один месяц) принять число рабочих дней в месяце равным 25-ти, а число рабочих часов в день, равным 8-ми.

Если случайные потоки, анализируемые в рассматриваемой задаче нельзя рассматривать как марковские, то необходимые характеристики СМО можно получить с помощью имитационного моделирования.