

Б.А.ЕСИПОВ

**СИСТЕМЫ И СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ**

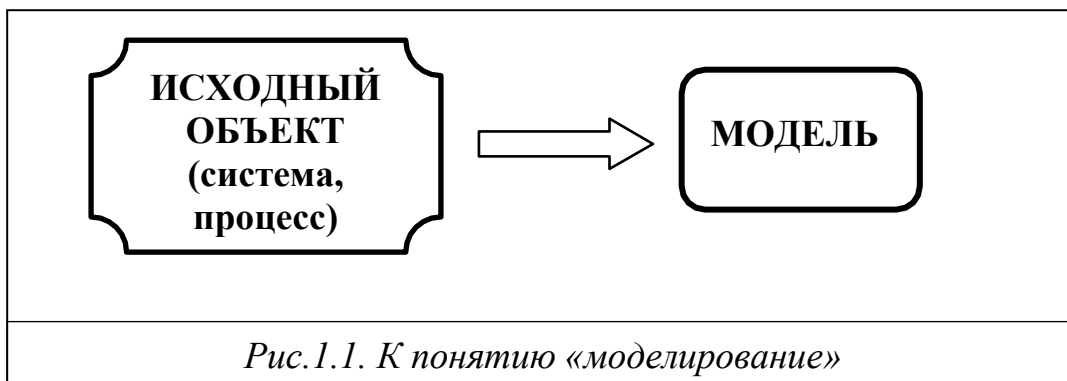
Курс лекций  
для магистров специальности ИВТ

Г.Самара  
2020г

## 1. ОБЩИЕ ВОПРОСЫ МОДЕЛИРОВАНИЯ

Изучение систем и сетей массового обслуживания (СМО и СеМО) базируется на методологии моделирования.

**Моделирование** – замещение одного *исходного объекта* другим объектом, называемым **моделью** (рис.1.1), и проведение экспериментов с моделью с целью получения информации о системе путем исследования свойств модели.



Объектами моделирования в технике являются *системы* и протекающие в них *процессы*. В частности, в вычислительной технике объектами моделирования являются вычислительные машины, комплексы, системы и сети. При этом, наибольший интерес представляют **конструктивные модели**, допускающие не только фиксацию свойств, но и исследование свойств систем (процессов) на построенной модели с целью решения задач проектирования систем с заданными свойствами.

Моделирование предоставляет возможность исследования таких объектов, прямой эксперимент с которыми:

- трудно выполним;
- экономически невыгоден;
- вообще невозможен.

Все более широкое распространение моделирование находит в таких областях как промышленные системы, экономика, социология, искусство, биология, медицина и т.п. В то же время, вычислительные системы и сети сами являются объектами моделирования на этапах проектирования новых и модернизации существующих систем, анализа эффективности использования систем в различных условиях (например, в экстремальных ситуациях, в условиях повышенных требований к надежности и живучести). Применение моделирования на этапе проектирования позволяет выполнить анализ различных вариантов предлагаемых проектных решений, определить работоспособность и оценить надежность системы, выявить узкие

места и мало загруженные ресурсы, а также сформулировать рекомендации по рациональному изменению состава и структуры или способа функциональной организации системы.

---

### 1.1. Свойства и виды моделей

**Модель** – физический или абстрактный объект, адекватно отображающий исследуемую систему.

Ко всем разрабатываемым моделям предъявляются два противоречивых требования:

- простота модели;
- адекватность исследуемой системе.

Требование *простоты модели* обусловлено необходимостью построения модели, которая может быть рассчитана доступными методами. Построение сложной модели может привести к невозможности получения конечного результата имеющимися средствами в приемлемые сроки и с требуемой точностью.

**Степень сложности (простоты)** модели определяется уровнем ее детализации, зависящим от принятых предположений и допущений: чем их больше, тем ниже уровень детализации и, следовательно, проще модель и, в то же время, менее адекватна исследуемой системе.

**Адекватность** (от лат. *adaequatus* – приравненный, равный) – соответствие модели оригиналу, характеризуемое степенью близости *свойств* модели свойствам исследуемой системы. Т.е. модель должна иметь такие же *свойства* как и оригинал, хотя физически они могут быть не похожи.

Адекватность математических моделей зависит от:

- степени полноты и достоверности сведений об исследуемой системе;
- уровня детализации модели.

При этом моделирование может проводиться:

- в условиях полной определенности, означающей наличие точной информации обо всех исходных параметрах;
- в условиях неопределенности, обусловленных:
  - неточностью сведений о параметрах;
  - отсутствием сведений о значениях некоторых параметров.

Многообразие систем, проявляющееся в многообразии их структурно-функциональной организации, определяет использование множества разных моделей, которые могут быть классифицированы в зависимости от:

1) *характера функционирования исследуемой системы:*

**детерминированные**, функционирование которых описывается детерминированными величинами;

**стохастические** или вероятностные, функционирование которых

описывается случайными величинами;

2) *характера протекающих в исследуемой системе процессов:*

- **непрерывные**, в которых процессы протекают непрерывно во времени;

- **дискретные**, в которых процессы меняют свое состояние скачкообразно в дискретные моменты времени;

3) *степени достоверности исходных данных об исследуемой системе:*

- с априорно известными параметрами;

- с неизвестными параметрами;

4) *режима функционирования системы:*

- **стационарные**, в которых характеристики не меняются со временем;

- **нестационарные**, в которых характеристики изменяются со временем;

5) *назначения:*

- **статические** или **структурные**, отображающие состав и структуру системы;

- **динамические** или **функциональные**, отображающие функционирование системы во времени;

- **структурно-функциональные**, отображающие структурные и функциональные особенности организации исследуемой системы;

б) *способа представления (описания) и реализации:*

- **концептуальные** или **содержательные**, представляющие собой описание (в простейшем случае словесное) наиболее существенных особенностей структурно-функциональной организации исследуемой системы;

- **физические** или **материальные** – модели, эквивалентные или подобные оригиналу (макеты) или процесс функционирования которых такой же, как у оригинала и имеет ту же или другую физическую природу;

- **математические** или **абстрактные**, представляющие собой формализованное описание системы с помощью абстрактного языка, в частности с помощью математических соотношений, отражающих процесс функционирования системы;

- **программные (алгоритмические, компьютерные)** – программы для ЭВМ, позволяющие наглядно представить исследуемый объект посредством имитации или графического отображения математических зависимостей, описывающих искомый объект.

Соответственно различают **физическое, математическое и компьютерное моделирование**.

Между классами систем и моделей обязательно должно существовать однозначное соответствие. Например, дискретные системы

---

могут быть представлены в виде непрерывных моделей, а детерминированные системы – в виде вероятностных моделей, и наоборот.

В дальнейшем основное внимание уделяется **математическому моделированию**, широко используемому при исследовании сложных технических и организационных систем, таких как системы и сети массового

обслуживания (СМО и СеМО).

## 1.2 Задачи моделирования

Моделирование, как процесс исследования сложных систем, в общем случае предполагает решение следующих взаимосвязанных задач:

- разработка модели;
- анализ характеристик системы;
- синтез системы;
- детальный анализ синтезированной системы.

Разработка модели состоит в выборе конкретного математического аппарата, в терминах которого формулируется модель, и построении модели или совокупности моделей исследуемой системы, отображающих возможные варианты структурно-функциональной организации системы. В процессе разработки модели необходимо определить состав и перечень параметров и характеристик модели в терминах выбранного математического аппарата, и установить их взаимосвязь с параметрами и характеристиками исследуемой системы, то есть выполнить *параметризацию* модели. Этот процесс иногда называют *калибровкой* модели.

Анализ характеристик системы с использованием разработанной модели заключается в выявлении свойств и закономерностей, присущих процессам, протекающим в системах с различной организацией, и выработке рекомендаций для решения основной задачи системного проектирования – задачи синтеза.

## 1.3. Методы моделирования

В зависимости от целей моделирование может проводиться на двух уровнях:

- *на качественном;*
- *на количественном.*

Математическое моделирование обычно проводится на количественном уровне с использованием конструктивных моделей.

При исследовании технических систем с дискретным характером функционирования наиболее широкое применение получили следующие методы математического моделирования:

- **аналитические** (аппарат теории вероятностей, теории массового обслуживания, теории случайных процессов, методы оптимизации, ...);
- **численные** (применение методов численного анализа для получения конечных результатов в числовой форме, когда невозможно получить аналитические зависимости характеристик от параметров в явном виде);

---

- **статистические** или **имитационные** (исследования на ЭВМ, базирующиеся на методе статистических испытаний и предполагающие применение специальных программных средств и языков моделирования: GPSS [4, 18], SIMULA, ИМСС [11] и др.).

- **комбинированные.**

#### 1.4. Аналитические методы

**Аналитические методы** состоят в построении математической модели в виде математических символов и отношений, при этом требуемые зависимости выводятся из математической модели последовательным применением математических правил.

*Достоинство* аналитических методов заключается в возможности получения решения в явной аналитической форме, позволяющей проводить детальный анализ процессов, протекающих в исследуемой системе, в широком диапазоне изменения параметров системы. Результаты в аналитической форме являются основой для выбора оптимальных вариантов структурно-функциональной организации системы на этапе синтеза.

*Недостаток* аналитических методов – использование целого ряда допущений и предположений в процессе построения математических моделей и невозможность, в некоторых случаях, получить решение в явном виде из-за неразрешимости уравнений в аналитической форме, (например, отсутствия первообразных для подынтегральных функций и т.п.). В этих случаях широко применяются численные методы.

Аналитические методы можно разделить на:

- точные;
- приближенные;
- эвристические.

#### 1.5. Численные методы

**Численные методы** основываются на построении конечной последовательности действий над числами. Применение численных методов сводится к замене математических операций и отношений соответствующими операциями над числами, например, к замене интегралов суммами, бесконечных сумм – конечными и т.п. Результатом применения численных методов являются таблицы и графики зависимостей, раскрывающих свойства объекта. Численные методы являются продолжением аналитических методов в тех случаях, когда результат не может быть получен в явном виде. Численные методы по сравнению с аналитическими методами позволяют решать значительно более широкий круг задач.

#### 1.6. Статистические методы

В тех случаях, когда анализ математической модели даже

численными методами может оказаться нерезультативным из-за чрезмерной трудоемкости или неустойчивости алгоритмов в отношении погрешностей аппроксимации и округления, строится имитационная модель, в которой процессы, протекающие в ВС, описываются как последовательности операций над числами, представляющими значения входов и выходов соответствующих элементов. *Имитационная модель* объединяет свойства отдельных элементов в единую систему. Производя вычисления, порождаемые имитационной моделью, можно на основе свойств отдельных элементов определить свойства всей системы.

При построении имитационных моделей широко используется *метод статистических испытаний* (метод Монте-Карло). Процедура построения и анализа имитационных моделей методом статистических испытаний называется **статистическим моделированием**. Статистическое моделирование представляет собой процесс получения статистических данных о свойствах моделируемой системы.

*Достоинством* статистического моделирования является *универсальность*, гарантирующая принципиальную возможность проведения анализа систем любой степени сложности с любой степенью детализации.

*Недостаток* статистического моделирования – *трудоемкость* процесса моделирования и *частный характер результатов*, не раскрывающий зависимости, а лишь определяющий ее в отдельных точках.

Статистическое моделирование широко используется для оценки погрешностей аналитических и численных методов.

### 1.7. Комбинированные методы

Комбинированные методы представляют собой комбинацию выше перечисленных методов, в частности:

- **численно-аналитические**, в которых часть результатов получается численно, а остальные – с использованием аналитических зависимостей;

- **аналитико-имитационные**, представляющие собой имитационное моделирование в сочетании с аналитическими методами, позволяющими сократить время моделирования за счет определения значений ряда характеристик на основе аналитических зависимостей по значениям одной или нескольких характеристик, найденных путем статистической обработки результатов имитационного моделирования.

**Рассмотренные методы моделирования применяются для разработки теории и практических методов исследования систем и сетей массового обслуживания (СМО и СеМО).**

## 2. Введение в системы и сети массового обслуживания (СМО и СеМО)

Во многих областях человеческой деятельности мы встречаемся с системами, в которых происходят характерные ситуации:

В них есть некоторые клиенты, которые требуют выполнения некоторой работы (*обслуживания*) внутри этой системы. Характерно то, что эти требования возникают в непредсказуемые моменты времени. Причем выполнение этой работы также длится непредсказуемое время. Это явление повторяется многократно – происходит массовое обслуживание.

*Примеры.* Ремонтные мастерские, бытовое обслуживание населения, магазины, очередь у кассы; автомобиль перед светофором; пробки на дорогах; самолет, требующий посадку (несмотря на расписание, время случайно); набор программ (задач), подлежащих обработке на компьютере.

Из-за случайного характера появления требований (заявок) и случайности времени обработки заявок, получаем случайные моменты изменений состояний системы.

В таких системах возникает очередь в обслуживании, клиент должен ждать, в других случаях система может простаивать.

Работоспособность такой системы как правило определяется средней величиной клиентов обслуживаемых в единицу времени, или средней скоростью обработки клиентов.

Не только эти величины определяют эффективность работы таких систем. Например важна вероятность возможного отказа в обслуживании и другие.

Сам характер поступления заявок в систему как правило характеризуется режимом "то густо, то пусто".

Системы такого рода называют системами массового обслуживания - СМО.

Если параметры обслуживания системы и потока клиентов был бы предсказуемым (детерминированным), то организовать работу системы можно было бы достаточно просто. Вся сложность организации работы СМО состоит в случайности процессов, протекающих в СМО.

Ясно, что основой для моделирования такой системы является теория вероятностей. Прикладная область этой науки, которая занимается исследованием процессов в СМО, выделилась в особое научное направление – теорию массового обслуживания ( ТМО ).

ТМО – устанавливает зависимости между характером потока заявок, количеством каналов обслуживания, их производительностью, правилами работы СМО и эффективностью обслуживания.

Взгляд на СМО двоякий:

– с точки зрения клиента: Например для клиента чем меньше он ждет, тем ему лучше. Простой системы ему так же на пользу.

– с точки зрения самой системы: важна пропускная способность системы, при этом ожидание в очереди клиентов может быть на пользу.



В задачу ТМО входит установление количественного отношения между отдельными параметрами системы, а затем на этой основе согласование характеристик владельцев и клиентов. Т.е. необходим разумный компромисс.

Зная математическую модель, терминах исследование операций можно какие-то параметры считать управляемыми, а какие-то – целевой функцией. И тогда можно поставить задачу оптимизации некоторых параметров СМО.

## 2. Основные компоненты модели СМО.

Само понятие СМО можно трактовать и как физическую реальную систему и как модель.

Рассмотрим виды упрощений, применяемых для моделирования СМО (парадигма модели СМО)

1) Все клиенты одинаковы – однородная масса.

2) Если заявка пришла, а обслуживающие приборы (каналы) заняты, она может встать в очередь, если очередь предусмотрена. т.е.:  
– существуют СМО с очередью (например, парикмахерская, бензозаправка)  
– существуют СМО без очереди (ожидания) (например, в проводном телефоне, в случае "занято" заявка покидает систему.)

3) Если очереди нет, то поступившей заявке отказывается в обслуживании. Т.е. существуют СМО с отказами.

Может быть, что очередь ограничена (число мест –  $m$ ), тогда если заявка пришла, и есть место в очереди – она в нее встает, иначе – «отказ».

В СМО принято, что все переходы из одного состояния в другое мгновенны, а каждое состояние – дискретная "картинка" (срез ситуации). Например в парикмахерской состояние: все три мастера заняты и два места из 10 стульев для ожидания заняты.

Каждому состоянию присуща вероятность этого состояния.

4) В СМО принято, что механизм обслуживания обобщается *одной* операцией. Но иногда вводят понятия фаза обслуживания – когда каждое обслуживание – это последовательность нескольких операций. Каждую фазу выполняет прибор. Иногда вводят фиктивные фазы обслуживания.

5) Все обслуживание сводится к регистрации времени обслуживания. Вот почему при моделировании в СМО рассматривают поток поступающих заявок (входной поток), а обслуженные заявки образуют выходной поток (поток обслуживаний). Эти потоки имеют различный характер. (Например, зашли с одной скоростью, встали в очередь, а скорость выхода зависит от обслуживания).

В общем случае входные и выходные потоки в СМО носят случайный характер и являются случайными процессами. Поэтому основными характеристиками входных и выходных потоков являются – распределение вероятностей количества заявок в единицу времени и распределение времени между заявками. Таким образом в СМО о случайных потоках мы знаем не больше, чем вероятностные распределения.

6) В некоторых СМО заявки могут обслуживаться группами (например, ресторан). Это система с параллельно-групповым обслуживанием.

7) Несмотря на то, что почти все определяется свойствами входных и выходных потоков, в СМО существуют и другие факторы, влияющие на работу:

1. Во многих СМО предусмотрена дисциплина управления очередью – правила, по которым из очереди выбирается заявка.

– FIFO – ПерППО, первый пришел, первый обслуживается

– LIFO – последний пришел, первый обслуживается.

– СОЗ – случайный отбор заявок.

2. **Приоритеты** в выборе заявок – когда предусмотрено несколько очередей (СМО с приоритетами)

3. Обслуживающая система имеет несколько приборов со своими каналами обслуживания, каждый канал может выполнять несколько фаз обслуживания. Возникает последовательность (тандем) очередей обслуживания. В этом случае возникает **сетевая структура СМО** -последовательно-параллельное обслуживание.

4. СМО характеризуется **допустимой величиной очереди** (буфера)  $m$ . Фактическая величина очереди в данный момент – это другая величина.

5. При анализе СМО необходимо учитывать факторы, связанные с природой источника заявок.

В ряде случаев заявки поступают «ниоткуда», и мы можем отвлечься от источника заявок. Это – **открытые СМО**.

В некоторых случаях приходится учитывать **емкость** источника, когда приходится считать количество заявок, которые могут выйти из источника. Это число может быть конечно. Кроме этого состояние обслуживающей системы может влиять на свойства источника заявок. Тогда сам источник включается в СМО. Это - **замкнутая СМО**.

*Пример.* В цехе 10 станков, которые обслуживают 2 ремонтника. Обслуживание станков. Источник заявок – неисправные станки. Когда они работают, они создают поток неисправностей. Емкость источника уменьшается, по мере увеличения неработающих станков.

6. Для СМО, в которой клиенты и обслуживаемы каналы – люди, приходится учитывать **бихевиоральные факторы**, связанные с поведением человеческих индивидуумов, например:

а) продавец способен повысить скорость обслуживания, когда очередь увеличивается

б) при параллельном обслуживании человек может переходить из очереди в очередь

в) человек может уйти, если очередь длинная.

При моделировании реальных СМО такие факторы нужно обобщить, т.к. в модели считается, что все клиенты одинаковы.

Таким образом, функциональная возможность любой модели массового обслуживания определяется следующими факторами:

1) Распределение моментов поступления заявок на обслуживание (единичных/групповых)

2) Распределение продолжительности обслуживания (при индивидуальном/групповом обслуживании)

3) Конфигурация обслуживания систем (последовательная, параллельная, последовательно-параллельная)

4) Дисциплина очереди (ПерППО, ПосППО, СОЗ) и приоритетные характеристики обслуживания системы

5) Вместительность блока ожидания (нет, ограничена, неограничена)

6) Емкость (мощность) источника заявок (конечна, бесконечно большая)

7) Бихевиоральные характеристики (переход из очереди в очередь, ненулевая вероятность отказа сразу, через некоторое время, увеличение скорости обслуживания при большой очереди)

Можно построить столько моделей, сколько существует комбинаций из этих факторов.

Современная ТМО рассматривает различные типы моделей, идя от простого к сложному.

Если можно смоделировать переход из всех состояний в другие и вычислить вероятности этих состояний, то можно учесть самые подробные факторы в СМО.

В настоящее время развиваются 2 пути анализа СМО:

1) **аналитический**, который позволяет получить аналитические (формульные) зависимости между параметрами СМО. Это возможно, если характер случайных потоков принадлежит к определенному классу распределений вероятностей.

2) Если же характер потоков не описывается классическими распределениями, а является результатом имитационных экспериментов, то характеристики СМО можно получить только как результат имитационного моделирования.

В настоящем курсе мы будем делать акцент на тех моделях, которые позволяют получить аналитические характеристики.

### 3. Модели теории массового обслуживания. (11.1)

ТМО базируется на применении математических моделей процессов, которые происходят в СМО. Вид и сложность моделей зависит от того, какие результаты хотим получить:

1) Модели для получения аналитических зависимостей, когда результирующие параметры СМО выражаются в виде формул, выраженные через параметры СМО и потоков, входящих в него. Для этого делается много упрощений.

2) Модели для получения численных итоговых результатов. Допускаются имитационные эксперименты, когда характер потоков моделируется согласно какой-то статистики. В современных компьютерных **инструментальных средствах имитационного моделирования** для анализа СМО присутствует множество специальных программ для упрощения взаимодействия человека, модели и самой СМО. Существуют специальные **языки имитационного моделирования СМО**.

В СМО происходят процессы случайного (стохастического) характера, а изучение процессов, протекающих в системах со стохастическим характером функционирования проводится в **теории случайных процессов** и ее приложении ТМО. ТМО иногда называют *теорией очередей*.

Все модели ТМО делятся на **базовые** в виде СМО и **сетевые** в виде СеМО. Это математические объекты, описываемые в терминах соответствующего математического аппарата.

## 4. Основные понятия ТМО

**I. СМО** – математический объект, который содержит 1 или несколько **каналов** к.

Этот канал обслуживает **заявки** З, поступающие в обрабатывающую систему и **накопитель** Н. В этом накопителе заявки образуют **очередь** О. Содержание заявки не интересно, интересуется время ее прихода. Совокупность заявок, распределенных во времени и поступающих в систему, образуют **поток заявок**.

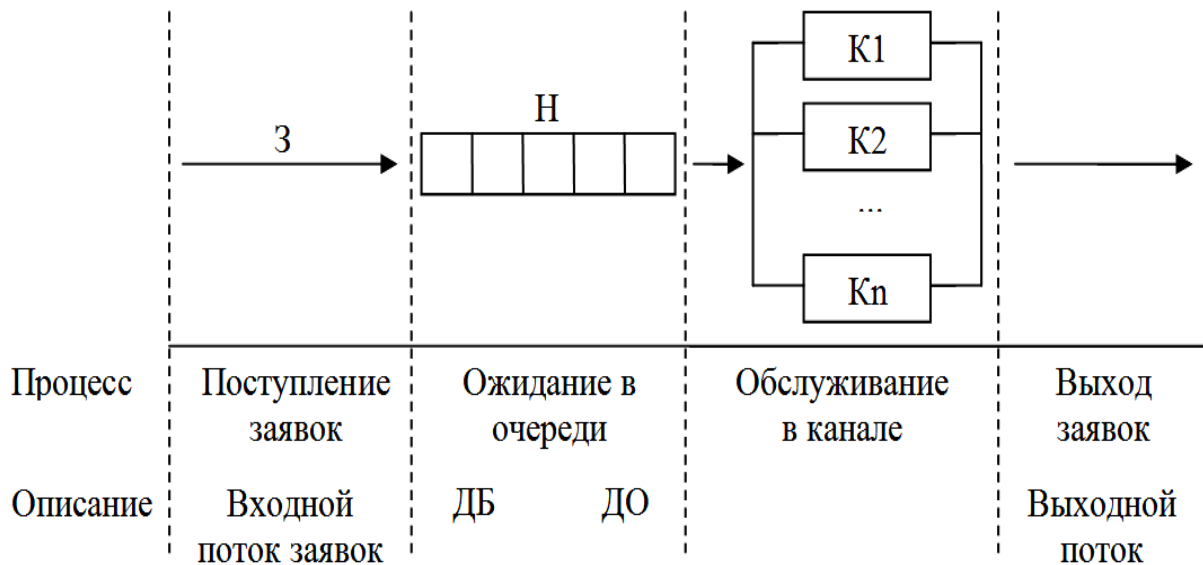
**Канал** (устройство) – элемент СМО, функционирование которого заключается в обслуживании заявок. В каждый момент времени в канале может быть только одна заявка.

**Обслуживание** – задержка заявки на некоторое время.

**Длительность обслуживания** – время задержки заявки в канале.

Обслуженная заявка выходит из системы и образует **поток обслуживания**.

**Накопитель** (буфер) – совокупность мест для ожидания заявок. Количество мест определяется **емкостью**  $m$ .



Количество заявок, ожидающих обслуживания в накопителе, называют **длиной очереди** ( $r$ ).

**Дисциплина буферизации (ДБ)** – правило занесения поступления заявок в накопитель.

**Дисциплина обслуживания (ДО)** – правило выбора заявок из очереди для обслуживания в канале.

**Приоритет** – преимущественные права занесения в накопитель или выбора из очереди для обслуживания в канале заявок одного класса.

Для аналитических методов расчета делается целый ряд **предположений (ограничений)**, которые несущественны во многих практических случаях. Но их применение позволяет получить аналитический результат.

В ТМО входят вопросы оценки того, как влияют ограничения (предположения) на результат анализа.

Упрощения:

- 1) В канале только 1 заявка
- 2) Заявка мгновенно попадает на обслуживание, если канал свободен
- 3) Из очереди в канал заявка попадает мгновенно
- 4) Поступление заявок в СМО и длительность их обслуживания не зависит от числа заявок в системе или от других факторов. (Это только для открытых СМО).
- 5) Длительность обслуживания не зависит от скорости поступления заявок

**II. СеМО** – совокупность взаимодействующих СМО, в среде которой циркулируют заявки.

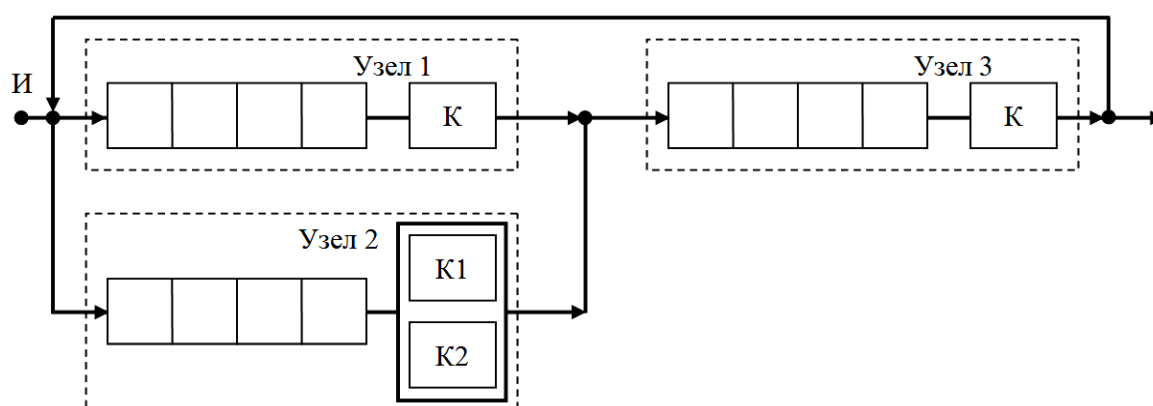
Дополнительным элементом в СеМО является узел У и источник заявок И.

Узел СеМО – СМО. Но когда рассматривается СеМО в виде графа, то внутри СМО не смотрят.

Источник – генератор заявок, поступающих в сеть и требующих определенных этапов обслуживания в узлах сети.

Математическая модель СеМО резко отличается от модели СМО.

Для упрощения изображения СеМО используется ориентированный граф. Дуги – переходы заявок. Вершины – узлы сети.



В моделях СеМО переходы заявок от узла к узлу задаются на основе вероятностных свойств работы самой сети. Для этого применяются определенные модели вероятностных распределений потоков.

Если моменты переходов дискретны, то потоки определяются вероятностью перехода, если же время непрерывно, то свойство потоков определяются плотностью распределения вероятностей между соседними событиями потоков.

Путь движения заявок в СеМО – маршрут.

## 5. Марковский случайны процесс. (11.2)

Для анализа дискретных систем применяется так называемый МЕТОД СОСТОЯНИЙ.

Состояние это – сложная «картинка» системы меняющаяся во времени. Работа системы – это цепь меняющихся состояний.

Приход заявки в СМО или СеМО меняет состояние системы.

Состояние обозначается буквой с индексами (номерах состояний), например  $Sn+r$  – состояние : n каналов занято и r заявок находятся в очереди. СМО - система с дискретными состояниями.

Процессы, происходящие в таких системах могут быть описаны аналитическими методами, если они принадлежат к марковским процессам.

### 1) 11.2. Марковский случайный процесс с дискретными состояниями

Определение: Случайный процесс, протекающий в системе S называется марковским, если для каждого момента времени  $t_0$  вероятность любого состояния системы в будущем (при  $t > t_0$ ) зависит только от ее состояния в настоящем (при  $t = t_0$ ) и не зависит от того, когда и каким образом система пришла в это состояние

(т.е. как процесс развивался в прошлом). На практике как правило свойство «марковости» выполняется. Поэтому марковские процессы широко применяются для моделирования реальных процессов.

Случайный процесс называется *процессом с дискретными состояниями*, если возможные состояния системы  $s_1, s_2, s_3, \dots$  можно перечислить, т.е. множество  $S$  счетное.

Кроме таких процессов имеются процессы с непрерывным изменением состояния, например, изменение электрического тока в лампе и т.п.

Будем изображать каждое состояние  $s_1, s_2, s_3, \dots$  прямоугольником, а возможные переходы за один шаг - стрелкой. Такие графические схемы называются *графами состояний*.

Пример.

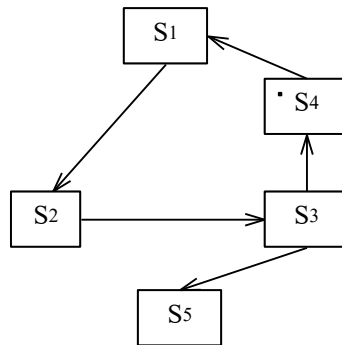


Рис. 69  $s_1$  - ЭВМ работает исправно;  $s_2$  - ЭВМ неисправна, ожидает осмотра;  $s_3$  - ЭВМ осматривается;  $s_4$  - ЭВМ ремонтируется;  $s_5$  - ЭВМ списана.

Случайный процесс называется *процессом с дискретным временем*, если переход системы из состояния в состояние возможен только в заранее фиксированные моменты времени  $t_1, t_2, \dots$

Случайный процесс называется *процессом с непрерывным временем*, если переход из одного состояния в другое возможен в любой момент времени.

Рассмотрим марковский случайный процесс с дискретными состояниями и дискретным временем. Известны

$$\left\{ \begin{array}{l} S_1, S_2, \dots, S_n \\ t_1, t_2, \dots, t_k \end{array} \right\}$$

Моменты  $t_1, t_2, \dots$  будем называть “шагами” или “этапами”. Случайный процесс заключается в том, что в последовательные моменты времени  $t_1, t_2, \dots$  система может быть в тех или иных состояниях из  $S_1, S_2, S_3, \dots$

Например:

$$S_1 \rightarrow S_3 \rightarrow S_4 \rightarrow S_2 \rightarrow S_{10} \rightarrow S_4 \rightarrow \dots$$

или

$$S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_4 \rightarrow S_1 \rightarrow \dots$$

Обозначим  $S_i^{(k)}$  - событие, состоящее в том, что после  $k$  шагов система находится в состоянии  $s_i$ . Тогда для любого  $k$  события  $S_1^{(k)}, S_2^{(k)}, S_3^{(k)}, \dots, S_n^{(k)}$  образуют полную группу событий.

Процесс можно еще записать так  $S_1^{(0)}, S_1^{(1)}, S_3^{(2)}, S_1^{(3)}, \dots$  Такая случайная последовательность называется *простой марковской цепью*, если для каждого

шага вероятность перехода из любого состояния  $S_i$  в любое  $S_j$  не зависит от того, когда и как система пришла в состояние  $S_i$ .

Опишем марковскую цепь при помощи вероятностей состояний. Тогда

$$P_1(1) = P\{S_1^{(1)}\}, P_2(1) = P\{S_2^{(1)}\} \text{ и т.д.,}$$

очевидно  $\sum_{i=1}^n P_i(k) = 1$  - условие полноты.

Будем называть  $P_i(k)$  - *вероятностями состояний*.

Найдем вероятности состояний системы для любого  $k$ . Марковскую цепь можно представить себе, как будто точка, изображающая систему  $S$ , случайным образом блуждает по графу состояний, “перескакивая” из состояния в состояние или задерживаясь в них.

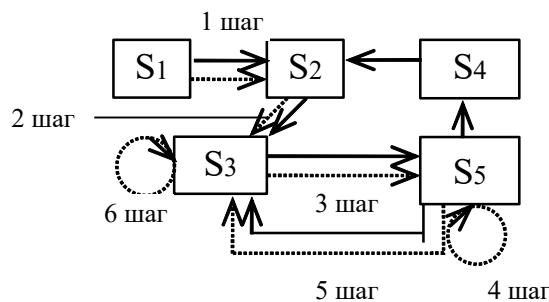


Рис. 70

(Пунктиром показаны действительные переходы, сплошной - возможные).

Для любого дискретного момента времени  $t_1, t_2, \dots$  существует вероятность перехода системы из одного состояния в другое (переходные вероятности).

Рассмотрим однородную марковскую цепь, имеющую  $n$  - состояний. Обозначим  $P_{ij}$  вероятность перехода из  $S_i$  в  $S_j$  ( $P_{ij}$  не зависит от  $k$ ).  $P_{ii}$  - вероятность задержки в состоянии  $S_i$ .

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix} \quad (11.1)$$

Существуют  $ij$ , для которых  $P_{ij} = 0$ . Используя события  $S_i^{(k)}$  можно  $P_{ij}$  записать в виде

$$P_{ij} = P\left\{ \frac{S_j^{(k)}}{S_i^{(k-1)}} \right\}$$

Из условия полноты сумма членов, стоящих в каждой строке матрицы (11.1) равна 1.

Марковская цепь называется *однородной*, если переходные вероятности не зависят от номера шага, в противном случае – *неоднородной*.

Имея размеченный граф состояний и зная начальное состояние системы, можно найти вероятности состояний  $P_1(k), P_2(k), \dots, P_n(k)$  после любого  $k$ -ого шага.

Легко показать (смотри рис. 71), что вероятности состояний  $P_i(k)$  после  $k$ -ого шага определяются рекуррентной формулой через вероятности состояний после  $(k-1)$  шага. Действительно, чтобы система была в момент  $k$  в состоянии  $S_i$  она должна быть в момент  $k-1$  в каком **либо** состоянии  $S_j$  **и** перейти с вероятностью  $P_{ji}$  в состояние  $S_i$ . Поэтому по формулам вероятности произведения и сложения совместных событий получаем формулу (11.2)

$$P_i(k) = \sum_{j=1}^n P_j(k-1)P_{ij}, \quad (i=1, \dots, n) \quad (11.2)$$

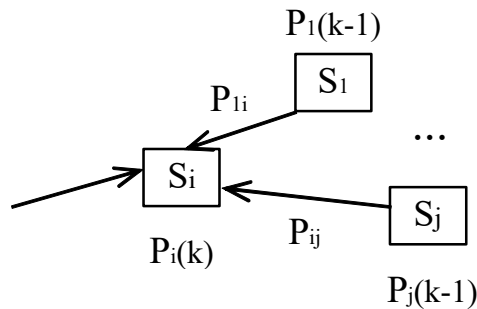


Рис. 71

Аналогично для неоднородной цепи Маркова

$$P_i(k) = \sum_{j=1}^n P_j(k-1)P_{ji}^{(k)}, \quad (i = \overline{1, n}), \quad (11.3)$$

где  $P_{ij}^k$  - вероятность перехода системы из состояния  $S_i$  в состояние  $S_j$  на  $k$ -м шаге

$$P_{ij}^{(k)} = P \left\{ \frac{S_j^{(k)}}{S_i^{(k-1)}} \right\}$$

## 2) 11.3. Марковский процесс с дискретным состоянием и непрерывным временем. Система уравнений Колмогорова

На практике часто переходы системы происходят в случайные моменты времени, которые заранее указать нельзя, Например, выход из строя аппаратуры, телефонный звонок и т.д.

Покажем, как выражаются вероятности состояний для этого процесса, который мы будем называть *непрерывной цепью Маркова*.

Оказывается, что все характеристики СМО легко получаются, если знать вероятности всех состояний СМО.

Пусть  $P_i(t)$  - вероятность того, что в момент  $t$  система  $S$  будет находиться в состоянии  $S_i$  ( $i=1, \dots, n$ ). Очевидно, для любого  $t$

$$\sum_{i=1}^n P_i(t) = 1, \quad \forall t$$



Здесь вместо переходных вероятностей  $P_{ij}$  нужно ввести плотность вероятностей перехода  $\lambda_{ij}$ .

Пусть в момент  $t$  система была в состоянии  $s_i$ . Рассмотрим  $\Delta t$ , примыкающий к  $t$ , тогда

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t}$$

где  $P_{ij}(\Delta t)$  - вероятность того, что систем за время  $\Delta t$  перейдем из  $s_i$  в  $s_j$  ( $i \neq j$ ). Очевидно

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

Если все плотности  $\lambda_{ij}$  не зависят от  $t$ , то марковский процесс называется *однородным (стационарным)*, в противном случае - *неоднородным (нестационарным)*.

Граф с проставленными у стрелок плотностями вероятностей перехода называется *размеченным графом состояний*.

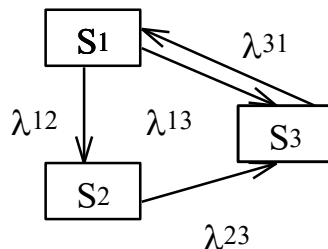


Рис. 72

Плотность вероятности перехода  $\lambda_{ij}$  еще называют *интенсивностью*, т.к. она показывает среднее количество переходов (событий) в единицу времени.

Пусть мы знаем размеченный граф состояний, тогда можно определить вероятности состояний  $P_1(t), P_2(t), \dots, P_n(t)$  как функции времени. Вероятности  $P_i(t)$  ( $i=1, 2, \dots, n$ ) удовлетворяют системе дифференциальных уравнений Колмогорова, решая которую, можно получить  $P_i(t)$ ,  $\forall i$ .

Покажем это на примере.

Возьмем четыре состояния  $S_1, S_2, S_3, S_4$ . Найдем  $P_1(t)$ .

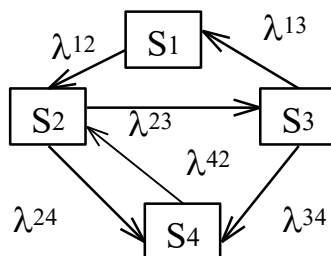


Рис. 73

Придадим  $t$  приращение  $\Delta t$  и найдем вероятность того, что в момент  $(t+\Delta t)$  система находится в  $s_1$ . Это событие может произойти двумя способами:

1. В момент  $t$  система была в состоянии  $s_1$ , а за время  $\Delta t$  не вышла из него.
2. В момент  $t$  система была в состоянии  $s_3$ , а за время  $\Delta t$  перешла из него в  $s_1$ .

Вероятность первого варианта, очевидно, будет найдена как произведение  $P_1(t)$  на вероятность того, что система не перейдет из  $S_1$

$$P_1(t) \cdot (1 - \lambda_{12}\Delta t).$$

Вероятность второго варианта

$$P_3(t) \cdot \lambda_{31}\Delta t$$

По правилу сложения вероятностей

$$P_1(t + \Delta t) = P_1(t)(1 - \lambda_{12}\Delta t) + P_3(t)\lambda_{31}\Delta t$$

Разделим на  $\Delta t$  и перейдем к пределу при  $\Delta t \rightarrow 0$ .

$$\begin{aligned} \frac{P_1(t + \Delta t) - P_1(t)}{\Delta t} &= -P_1(t)\lambda_{12} + P_3(t)\lambda_{31} \\ \lim_{\Delta t \rightarrow 0} \frac{P_1(t + \Delta t) - P_1(t)}{\Delta t} &= \lim_{\Delta t \rightarrow 0} (-P_1(t)\lambda_{12} + P_3(t)\lambda_{31}) \\ \frac{dP_1(t)}{dt} &= -P_1(t)\lambda_{12} + P_3(t)\lambda_{31} \end{aligned}$$

Рассмотрим  $S_2$  и найдем  $P_2(t + \Delta t)$  - вероятность того, что в момент  $(t + \Delta t)$  система будет находиться в  $S_2$ . Здесь это событие может произойти тремя способами:

1. В момент  $t$  система была в состоянии  $S_2$ , а за время  $\Delta t$  не вышла из него.
2. В момент  $t$  система была в состоянии  $S_1$ , а за время  $\Delta t$  перешла из него в  $S_2$ .
3. В момент  $t$  система была в состоянии  $S_4$ , а за время  $\Delta t$  перешла из него в  $S_2$ .

$$\text{Для (1): } P_2(t) \cdot [1 - \lambda_{23}\Delta t - \lambda_{24}\Delta t],$$

$$\text{для (2): } P_1(t) \cdot \lambda_{12}\Delta t,$$

$$\text{для (3): } P_4(t) \cdot \lambda_{42}\Delta t$$

$$P_2(t + \Delta t) = P_2(t)(1 - \lambda_{23}\Delta t - \lambda_{24}\Delta t) + P_1(t)\lambda_{12}\Delta t + P_4(t)\lambda_{42}\Delta t$$

$$\frac{dP_2(t)}{dt} = -P_2(t)\lambda_{23} - P_2(t)\lambda_{24} + P_1(t)\lambda_{12} + P_4(t)\lambda_{42}$$

Аналогично для  $S_3, S_4$  получаем

$$\left. \begin{aligned} \frac{dP_1}{dt} &= -\lambda_{12}P_1 + \lambda_{31}P_3 \\ \frac{dP_2}{dt} &= -\lambda_{23}P_2 - \lambda_{24}P_2 + \lambda_{12}P_1 + \lambda_{42}P_4 \\ \frac{dP_3}{dt} &= -\lambda_{31}P_3 - \lambda_{34}P_3 + \lambda_{23}P_2 \\ \frac{dP_4}{dt} &= -\lambda_{42}P_4 + \lambda_{24}P_2 + \lambda_{34}P_3 \end{aligned} \right\} \quad (11.4)$$

Это и есть система уравнений Колмогорова для вероятностей состояний. Интегрируя их, можно найти  $P_i(t)$ ,  $\forall i$ . Начальные условия берутся в зависимости от начального состояния системы. Если при  $t = 0$  система была в  $S_1$ , то надо принять

при  $t=0$ :  $P_1=1; P_2=P_3=P_4=0$ .

Заметим, что в системе (11.4) достаточно трех уравнений, т.к.

$$P_4 = 1 - (P_1 + P_2 + P_3).$$

Структура (11.4) имеет закономерность, которая позволяет записать систему уравнений Колмогорова по виду графа состояний.

Алгоритм составления уравнений Колмогорова:

В левой части каждого уравнения стоит производная вероятности состояния, а правая часть содержит столько членов, сколько стрелок связано с данным состоянием. Если стрелка направлена из состояния, то соответствующий член имеет знак “-”, если в состояние - знак “+”. Каждый член равен произведению плотности вероятности перехода, соответствующей данной стрелке, умноженной на вероятность того состояния, из которого исходит стрелка.

Эти правило справедливо для любых марковских цепей и позволяет механически записывать уравнения Колмогорова.

Пример.

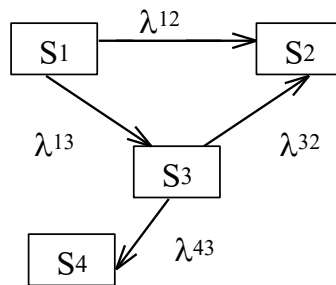


Рис. 74

$$t = 0; P_1 = 1; P_2 = P_3 = P_4 = 0$$

$$\frac{dP_1}{dt} = -(\lambda_{12} + \lambda_{13})P_1$$

$$\frac{dP_2}{dt} = P_1\lambda_{12} + \lambda_{32}P_3$$

$$\frac{dP_3}{dt} = -\lambda_{32}P_3 + \lambda_{13}P_1 - \lambda_{43}P_3$$

$$\frac{dP_4}{dt} = \lambda_{43}P_3$$

### 3) 11.4. Поток событий. Простейший поток и его свойства

*Поток событий* - последовательность однородных событий, следующих одно за другим в случайные моменты времени.

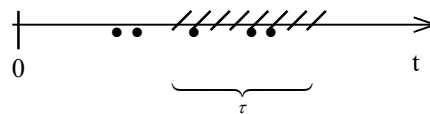


Рис.75

1. Поток событий называется *стационарным*, если вероятность попадания того или иного числа событий на участок времени длиной  $\tau$  зависит только от длины участка и не зависит от местонахождения этого участка.

(Стационарность может быть локальной, когда указанное свойство выполняется только на каком-то интервале времени.)

2. Поток событий называется потоком *без последствия*, если для любых непересекающихся участков времени число событий, попадающих на один из них, не зависит от того, сколько событий попало на другой (другие).
3. Поток называется *ординарным*, если вероятность попадания на элементарный участок двух или более событий пренебрежимо мала по сравнению с вероятностью попадания ординарных событий. (Например, поток в ЗАГС не является ординарным).

Поток, удовлетворяющий свойствам (1), (2), (3) называется *простейшим пуассоновским потоком*.

Можно доказать, что при взаимном наложении достаточно большого числа потоков, обладающих последствием (лишь бы они были стационарными и однородными) образуется простейший поток.

Если поток не имеет последствия, ординарен, но не стационарен, он называется *нестационарным пуассоновским потоком*. В таком потоке *интенсивность*  $\lambda$  (среднее число событий в единицу времени) зависит от времени

$$\lambda = \lambda(t)$$

Пуассоновский (стационарный и нестационарный) поток тесно связан с известным распределением Пуассона, в соответствии с которым число событий потока, попадающих на любой участок, распределено по закону Пуассона.

Рассмотрим участок от  $t_0$  до  $t_0 + \tau$ , тогда как известно из теории вероятностей, вероятность попадания на  $\tau$   $m$  событий

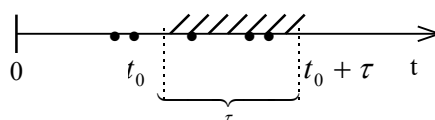


Рис. 76

$$P_m = \frac{a^m}{m!} e^{-a}, \quad (m=0,1,\dots),$$

где  $a$  - среднее число событий, приходящееся на участок  $\tau$ .

Для стационарного пуассоновского потока

$$a = \lambda \tau,$$

т.е. не зависит от того, где находится участок  $\tau$ .

Для нестационарного пуассоновского потока

$$a = \int_{t_0}^{t_0 + \tau} \lambda(t) dt$$

В теории вероятностей доказано, что выполнение свойств 1, 2, 3 приводит к пуассоновскому закону распределения, т.о. простейший поток всегда пуассоновский.

Рассмотрим простейший поток событий с интенсивностью  $\lambda$ . Найдем закон распределения  $T$  - *времени между событиями*.

Сначала найдем интегральный закон

$$F(t) = P(T < t)$$

( $T < t$ ) будет тогда, когда в участок длины  $t$  попадет хотя бы одно событие

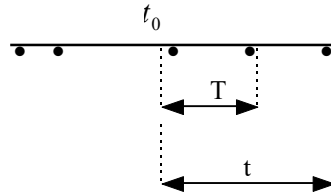


Рис. 77

$$F(t) = 1 - P_0$$

где  $P_0$  - вероятность того, что не попадет ни одного события

$$P_0 = \frac{a^0}{0!} e^{-a} = e^{-a} = e^{-\lambda t}$$

$$F(t) = 1 - e^{-\lambda t}, (t > 0)$$

$$f(t) = \frac{\partial F(t)}{\partial t} = \lambda e^{-\lambda t}, (t > 0) \quad (11.5)$$

Этот закон распределения называется *показательным*.  $\lambda$  - показатель *показательного закона*.

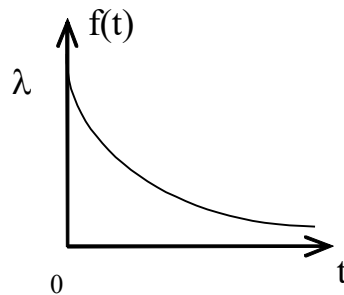


Рис. 78

$$m_t = \int_0^{\infty} t f(t) dt = \lambda \int_0^{\infty} t e^{-\lambda t} dt = \frac{1}{\lambda},$$

$$D_t = \int_0^{\infty} t^2 f(t) dt - m_t^2 = \frac{1}{\lambda^2}, \quad (11.6)$$

$$\sigma_t = \frac{1}{\lambda},$$

$$m_t = \sigma_t = \frac{1}{\lambda}.$$

Последнее равенство используется для проверки того, является ли поток пуассоновским. На рис.79 показан вид функции плотности  $f(t)$  при различных  $\lambda$ .

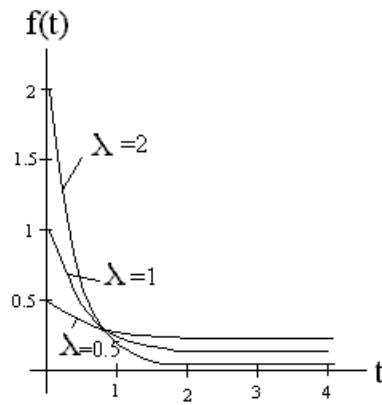


Рис 79

Вычислим вероятность появления интервалов, меньших математического ожидания  $\frac{1}{\lambda}$ .

$$P\left\{t < \frac{1}{\lambda}\right\} = \int_0^{\frac{1}{\lambda}} \lambda \cdot e^{-\lambda t} dt = 1 - \frac{1}{e} = 0.63$$

Т.е. в простейшем пуассоновском потоке короткие интервалы более часты, чем длинные, что создает более тяжелые условия работы, т.к. заявки группируются.

Для нестационарного потока закон распределения промежутка  $T$  уже не будет показательным, т.к. закон зависит от вида функции  $\lambda(t)$ .

Найдем вероятность того, что на участке  $\Delta t$  появится какое-то событие потока, т.е. участок не будет пустым. Так как поток ординарен, то

$$P_1(\Delta t) \approx 1 - P_0(\Delta t)$$

где  $P_1(\Delta t)$  - вероятность того, что на  $\Delta t$  появится ровно одно событие;  $P_0(\Delta t)$  - ни одного события

$$P_0(\Delta t) = \frac{a^0}{0!} e^{-a} = e^{-a} = e^{-\lambda \Delta t}$$

$$P_1(\Delta t) \approx 1 - e^{-\lambda \Delta t}$$

Разложим  $e^{-\lambda \Delta t}$  в ряд

$$P_1(\Delta t) \approx 1 - (1 - \lambda \Delta t)$$

$$P_1(\Delta t) \approx \lambda \Delta t$$

Для нестационарного потока

$$P_1(\Delta t) \approx \lambda(t) \Delta t$$

Полученные формулы позволяют очень просто находить переходные вероятности (плотности) для составления уравнений Колмогорова (см. п. 11.3.)

Кроме пуассоновских потоков в теории массового обслуживания рассматриваются потоки Пальма и Эрланга, являющиеся обобщением пуассоновских потоков.

*Поток Пальма* - это поток, в котором промежутки времени между последовательными событиями  $T_i$ , есть независимые и одинаково распределенные случайные величины, т.о. если они распределены по показательному закону, то поток Пальма - пуассоновский поток.

*Поток Эрланга  $k$  порядка* получается из простейшего, если в простейшем сохранить каждую  $k$  точку, а остальные выбросить. Простейший поток - это поток Эрланга 1 порядка.

В дальнейшем, рассматривая марковские процессы в системах с дискретными состояниями и непрерывным временем (непрерывные марковские цепи), нам удобно будет всегда рассматривать переходы системы из состояния в состояние как происходящие под влиянием каких-то потоков событий.

Пусть рассматривается система  $S$ , в которой переходы из состояния в состояние происходят под действием пуассоновских потоков с определенными интенсивностями (эти интенсивности ставят на графе состояний).

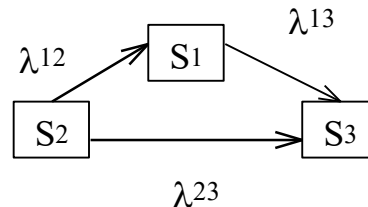


Рис. 80

После этого можно составить систему уравнения Колмогорова, считая, что

$$P_1(\Delta t) \approx \lambda \Delta t$$

т.е.  $\lambda_{ij}$  играет здесь роль плотностей.

Система уравнения Колмогорова - это система линейных дифференциальных уравнений с постоянными (или с переменными, если  $\lambda = \lambda(t)$ ) коэффициентами. Такие системы легко решаются с использованием ЭВМ.

#### 4) 11.5. Предельные вероятности состояний

Для практических расчетов нет необходимости знать функции  $P_i(t)$ . Достаточно знать так называемые *предельные вероятности*  $P_i$ .

Пусть имеется система  $S$  с дискретными состояниями  $S_1, S_2, \dots, S_n$ , в которых протекает марковский случайный процесс с непрерывным временем. Пусть все интенсивности постоянны

$$\lambda_{ij} = \text{const.}$$

Записав и проинтегрировав систему уравнений Колмогорова, мы получим вероятности состояний как функции времени

$$P_1(t), P_2(t), \dots, P_n(t)$$

$$\forall t, \quad \sum_{i=1}^n P_i(t) = 1$$

Возникает вопрос, что будет с  $P_i(t)$  при  $t \rightarrow \infty$ .

Пределы  $\lim_{t \rightarrow \infty} P_i(t) = P_i$  (если они существуют) называются *предельными вероятностями состояний*.

Можно доказать следующее положение:

если число состояний системы конечно, из каждого состояния можно перейти (за конечное число шагов) в другое, то предельные вероятности состояний существуют и не зависят от начального состояния системы.

Условие выполняется:

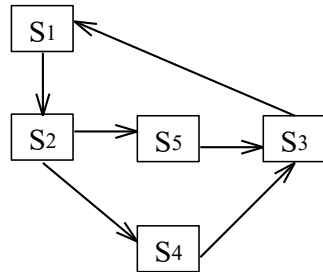


Рис. 81

Условие не выполняется:

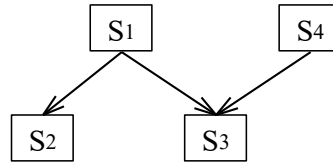


Рис. 82

Очевидно для предельных вероятностей должно быть:

$$\sum_{i=1}^n P_i = 1$$

т.е. при  $t \rightarrow \infty$  в системе устанавливается предельный стационарный режим. В этом режиме  $P_i(t) = P_i$  не зависят от времени - это среднее относительное время системы в данном состоянии.

Если  $P(S_1) = 0.2; P(S_2) = 0.3; P(S_3) = 0.5$ ,

20% времени система находилась в состоянии  $S_1$ ,

30% времени система находилась в состоянии  $S_2$ ,

50% времени система находилась в состоянии  $S_3$

Как же вычислить предельные вероятности состояний?

Оказывается для этого надо в уравнениях Колмогорова положить все левые части (производные) равными нулю, а  $P_i$  считать постоянными.

Действительно, в установившемся режиме  $\frac{dP_i}{dt} = 0$ , отсюда получим систему линейных алгебраических уравнений с нормировочными условием

$$\sum_{i=1}^n P_i = 1$$

Решив ее, найдем все предельные вероятности состояний. Покажем это для практически важных примеров моделей СМО.

## 5) 11.6. Процесс гибели и размножения

Марковская непрерывная цепь называется “процессом гибели и размножения”, если граф состояний имеет вид, представленный на рисунке.

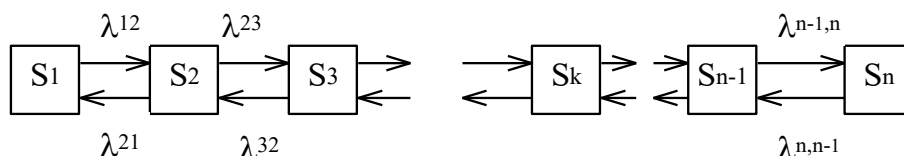


Рис. 83



Например: Каждый узел технической системы может выходить из строя и ремонтироваться. Тогда  $s_i$  - состояние, когда  $(i-1)$  узлов вышло из строя.

Напишем систему алгебраических уравнений для предельных вероятностей состояний.

$$\text{Для } S_1 \quad \lambda_{12}P_1 = \lambda_{21}P_2,$$

$$\text{для } S_2 \quad \lambda_{23}P_2 + \lambda_{21}P_2 = \lambda_{12}P_1 + \lambda_{32}P_3.$$

Первое равенство можно сократить, получим

$$\lambda_{23}P_2 = \lambda_{32}P_3$$

.....

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

$$\forall (k = 2 \dots n), \begin{cases} \lambda_{k-1,k}P_{k-1} = \lambda_{k,k-1}P_k \\ P_1 + P_2 + \dots + P_n = 1 \end{cases} \text{ получим}$$

Решая последовательно, получаем

$$P_2 = \frac{\lambda_{12}}{\lambda_{21}} P_1;$$

$$P_3 = \frac{\lambda_{23}}{\lambda_{32}} P_2 = \frac{\lambda_{23}}{\lambda_{32}} \cdot \frac{\lambda_{12}}{\lambda_{21}} P_1 \quad (11.7)$$

$$P_k = \frac{\lambda_{k-1,k} \lambda_{k-2,k-1} \dots \lambda_{12}}{\lambda_{k,k-1} \lambda_{k-1,k-2} \dots \lambda_{21}} P_1, \quad k = 2, \dots, n$$

т.о. все  $P_i$  выражены через  $P_1$ . Подставим их в полученные условие

$$\sum_{i=1}^n P_i = 1$$

$$\text{получим } P_1 = \frac{1}{1 + \sum_{k=2}^n \frac{\lambda_{k-1,k} \lambda_{k-2,k-1} \dots \lambda_{12}}{\lambda_{k,k-1} \lambda_{k-1,k-2} \dots \lambda_{21}}} \quad (11.8)$$

остальные  $P_i$  выражаются через  $P_1$ .

## 6) 11.7. Циклический процесс

Циклический процесс имеет граф состояний рис.84

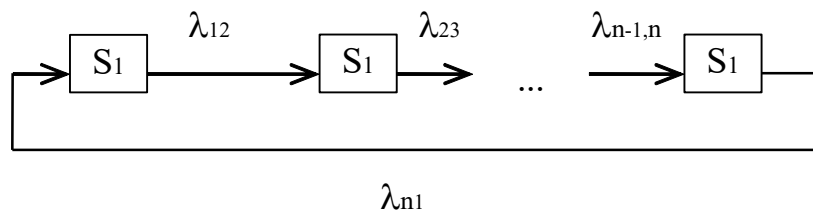


Рис. 84

Выведем предельные вероятности. Система алгебраических уравнений Колмогорова будет такой:

$$\left. \begin{array}{l} \lambda_{23}P_2 = \lambda_{12}P_1 \\ \lambda_{34}P_3 = \lambda_{23}P_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \\ \lambda_{k-1,k}P_{k-1} = \lambda_{k,k-1}P_k \\ \cdot \quad \cdot \quad \cdot \quad \cdot \\ \lambda_{n1}P_n = \lambda_{12}P_1 \\ \sum_{i=1}^n P_i = 1 \end{array} \right\} \begin{array}{l} P_2 = \frac{\lambda_{12}}{\lambda_{23}} P_1 \\ P_3 = \frac{\lambda_{23}}{\lambda_{34}} P_2 = \frac{\lambda_{12}}{\lambda_{23}} \frac{\lambda_{23}}{\lambda_{34}} P_1 = \frac{\lambda_{12}}{\lambda_{34}} P_1 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \\ P_k = \frac{\lambda_{12}}{\lambda_{k,k-1}} P_1 \end{array} \quad (11.9)$$

Подставляя в  $\sum_{i=1}^n P_i = 1$  получаем

$$P_1 = \frac{1}{1 + \lambda_{12} \left( \frac{1}{\lambda_{23}} + \frac{1}{\lambda_{34}} + \dots + \frac{1}{\lambda_{n1}} \right)} \quad (11.10)$$

и получаем  $P_k, \forall (k = 2 \dots n)$ .

Перейдем от интенсивностей  $\lambda_{ij}$  к средним временам  $\bar{t}_i$  пребывания системы в состоянии  $S_i$ .

Из  $S_i$  в циклической схеме исходит одна стрелка. Найдем математическое ожидание времени  $T_i$ , в течение которого система пробудет в состоянии  $S_i$ . Т.к. процесс марковский, закон распределения  $T_i$  не зависит от того, сколько времени система уже пробыла в состоянии  $S_i$ , т.е. закон такой же, если бы система только что пришла в состояние  $S_i$ . Поэтому промежуток  $T$  между событиями имеет показательный закон распределения. Параметр его  $\lambda_{i,i+1}$ , тогда

$$\bar{t}_i = \frac{1}{\lambda_{i,i+1}},$$

$$\lambda_{i,i+1} = \frac{1}{\bar{t}_i}, \forall i.$$

В силу цикличности  $\lambda_{n1} = \frac{1}{\bar{t}_n}$ , подставив это в полученные формулы для вероятностей получим

$$P_k = \frac{\bar{t}_k}{\sum_{i=1}^n \bar{t}_i}, \quad (k = 1 \dots n), \quad (11.11)$$

Т.е. предельные вероятности состояний в циклической схеме относятся как средние времена пребывания системы подряд в каждом из состояний.

Например: граф состояний для СМО – ЭВМ.

$$P_1 = 12 / (12 + 0,5 + 6 + 1) = 12 / 19,5;$$

$$P_2 = 1/19,5;$$

$$P_3=6/19,5; \quad P_4=0,5/19,5.$$

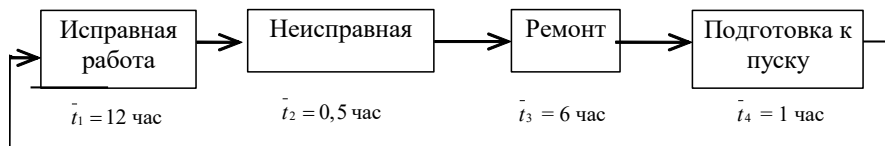


Рис. 85

## 7) 11.8. Классификация СМО

Различают два основных типа СМО.

1. Системы с отказами. Здесь заявки, попадающие в момент, когда все каналы заняты, получают “отказ” и покидают СМО, и в обслуживании далее не участвуют.

2. Системы с ожиданием (с очередью). Заявки, в случае занятости каналов, становятся в очередь и ожидают, пока не освободится какой-либо канал. Затем после освобождения какая-либо заявка обслуживается (не обязательно первоочередная).

Системы с ожиданием бывают *упорядоченные* (т.е. обслуживание происходит в порядке очереди) и *неупорядоченные* (в случайном порядке). Бывают СМО с *приоритетом*, когда более важные заявки обслуживаются в первую очередь.

СМО с ожиданием делятся на системы с *неограниченным ожиданием* и с *ограниченным ожиданием*. В первой любые заявки, поступившие в СМО, будут рано или поздно обслужены, во второй есть предел для ожидания.

В зависимости от типов СМО применяются различные критерии эффективности их работы:

1. В СМО с отказами:

- *абсолютная пропускная способность* - среднее число заявок, которые может обслужить СМО за единицу времени  $Q$ ;
- *относительная пропускная способность СМО* - средняя доля поступивших заявок, обслуживаемая системой

$$q = \frac{\text{обслуженные}}{\text{все заявки (обслуженные + необслуженные)}}$$

Также рассматривается *среднее число занятых каналов и среднее относительное время простоя системы и каналов*.

2. Для СМО с неограниченным ожиданием абсолютная и относительная пропускная способность теряет смысл, т.к. любая заявка когда-нибудь все равно обслуживается. Вместо них используются:

- среднее число заявок в очереди;
- среднее число заявок в системе (в очереди + обслуживаемые);
- среднее время ожидания в очереди;
- среднее время пребывания заявки в системе и др.

Для СМО с ограниченным ожиданием все вышеуказанные характеристики могут быть использованы.

Подводя итог описанию процессов, происходящих в СМО, можно сказать, что для анализа процесса в СМО нужно знать:

- $n$  - число каналов;

- - интенсивность потока заявок;
- - производительность каналов (среднее число обслуживаемых в единицу времени заявок);
- условия образования очереди.

Используем полученные формулы для анализа конкретных СМО. При этом все потоки считаем простейшими.

### 8) 11.9. Одноканальная СМО с отказами

Это простейшая СМО  $n=1$ . Входной поток пуассоновский с интенсивностью  $\lambda$ . Плотность распределения времени между соседними событиями входного потока имеет экспоненциальный вид

$$f(t) = \lambda e^{-\lambda t}, (t > 0),$$

со средним значением  $1/\lambda$  и дисперсией  $1/\lambda^2$ .

Заявка, в случае занятости канала, покидает СМО.

Обслуживание заявки, в общем случае, продолжается в течение случайного времени обслуживания  $T_{об}$ , распределенного также по показательному закону с параметром  $\mu$ .

$$f(t) = \mu e^{-\mu t}, \quad t > 0,$$

со средним временем обслуживания  $1/\mu$ .

Т.о. поток обслуживаний также простейший.

Требуется найти:

1. Абсолютную пропускную способность СМО  $Q$ .
2. Относительную пропускную способность  $q$ .

Канал обслуживания может находиться в двух состояниях  $S_0$  – свободен,  $S_1$  – занят.

Граф состояний будет таким

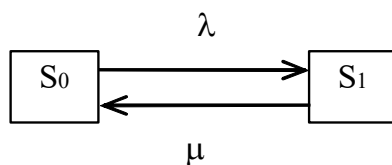


Рис. 86

Очевидно, что

$$P_0(t) + P_1(t) = 1;$$

Система дифференциальных уравнений Колмогорова запишется так

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

$$\frac{dP_1}{dt} = -(\mu + \lambda)P_1 + \lambda P_0.$$

Пусть  $P_0(0)=1$ ;  $P_1(0)=1$ ; т.о. в  $t=0$  заявок нет.

Решим дифференциальное уравнение аналитически

$$P_0 = uv$$

$$u'v + v'u = -(\mu + \lambda)uv + \mu;$$

$$v[u' + (\mu + \lambda)u] + v'u - \mu = 0;$$

$$u' + (\mu + \lambda)u = 0;$$

$$\frac{du}{dt} = -(\mu + \lambda)u;$$

$$\frac{du}{u} = -(\mu + \lambda)dt;$$

$$\ln u = -(\mu + \lambda)t$$

$$u = e^{-(\mu + \lambda)t}$$

$$v'e^{-(\mu + \lambda)t} = \mu$$

$$dv = \mu e^{(\mu + \lambda)t} dt$$

$$v = \mu \frac{1}{\mu + \lambda} e^{(\mu + \lambda)t} + c$$

$$P_0 = e^{-(\mu + \lambda)t} \left( \mu \frac{1}{\mu + \lambda} e^{(\mu + \lambda)t} + c \right); \text{ Н.у. } P_0(0) = 1$$

$$1 = c + \frac{\mu}{\mu + \lambda}; c = \frac{\lambda}{\mu + \lambda}$$

Окончательно получаем искомые вероятности как функции времени в виде

$$P_0 = \frac{\mu}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} e^{-(\mu + \lambda)t}$$

$$P_1 = 1 - P_0 = \frac{\lambda}{\mu + \lambda} + \frac{\mu}{\mu + \lambda} e^{-(\mu + \lambda)t}$$

Эта зависимость имеет вид:

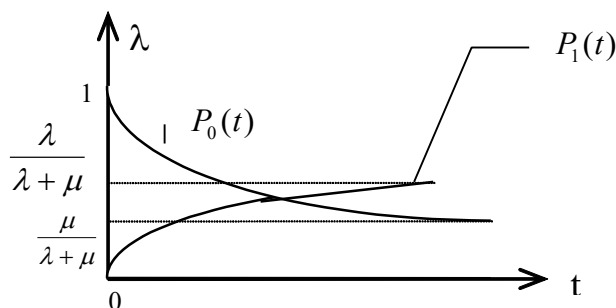


Рис. 87

Из графика видно, что предельные вероятности  $P_0$  и  $P_1$  - это значения, к которым стремятся  $P_0(t)$  и  $P_1(t)$ , т.о.

$$P_0 = \frac{\mu}{\mu + \lambda}; \quad P_1 = \frac{\lambda}{\mu + \lambda}$$

Очевидно, предельные вероятности можно получить, если сразу составить систему алгебраических уравнений Колмогорова.

$P_0$  - есть вероятность того, что заявка, пришедшие в момент  $t$ , будет обслужена, следовательно, это и есть относительная пропускная способность

$$q = P_0$$

$$q = \frac{\mu}{\lambda + \mu};$$

Вероятность отказа – это в данном случае  $P_1$ .

Т.к. только  $q$ -я доля из пришедших  $\lambda$  заявок обслуживается системой, то абсолютная пропускная способность  $Q$  получается в виде

$$Q = \lambda q$$

$$Q = \frac{\lambda \mu}{\lambda + \mu};$$

Очевидно, что вероятность отказа будет

$$P_{отк} = 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu}$$

## 9) 11.10. Многоканальная СМО с отказами

Пусть число каналов СМО равно  $n$ .

Перечислим все состояния СМО.

$S_0$  – все каналы свободны;

$S_1$  – занят один канал;

... ..

$S_k$  – занято  $k$  каналов,  $(n - k)$  – свободны;

... ..

$S_n$  – заняты все  $n$  каналов.

Граф состояний будет иметь вид:

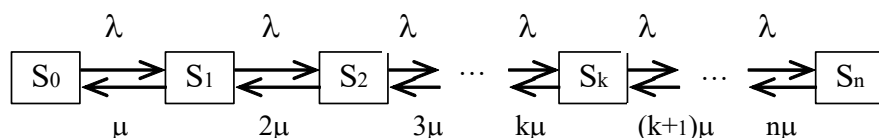


Рис. 88

По стрелкам слева направо систему переводит один и тот же поток с интенсивностью  $\lambda$ . Если система была в  $S_k$  (занято  $k$  каналов), то любая заявка переводит ее в  $S_{k+1}$ .

По стрелкам справа налево:  $S_1 \rightarrow S_0$  после окончания обслуживания одного канала система перейдет в  $S_0$ . Если занято  $k$  каналов, то общий поток обслуживаний будет в “ $k$ ” раз интенсивней.

Это частный случай процесса “гибели и размножения” (см. п. 11.6).

Составим систему уравнений Колмогорова:

$$\left. \begin{aligned} \frac{dP_0}{dt} &= -\lambda P_0 + \mu P_1 \\ \frac{dP_1}{dt} &= -(\lambda + \mu)P_1 + \lambda P_0 + 2\mu P_1 \\ &\dots \\ \frac{dP_k}{dt} &= -(\lambda + k\mu)P_k + \lambda P_{k-1} + (k+1)\mu P_{k+1} \\ &\dots \\ \frac{dP_n}{dt} &= -n\mu P_n + \lambda P_{n-1} \end{aligned} \right\} \quad (11.12)$$

(11.12) называются уравнениями Эрланга.

Начальные условия:  $P_0(0)=1$ ;  $P_1(0)=P_2(0)=\dots=P_n(0)=0$ . В начальный момент система свободна. Решив (10.12) найдем  $P_k(t), \forall k$ .

Предельные вероятности найдем, используя формулы для случая процесса гибели и размножения (11.7), (11.8).

$$P_k = \frac{\lambda^k \cdot P_0}{\mu \cdot 2\mu \dots k\mu} = \left(\frac{\lambda}{\mu}\right)^k \cdot \frac{P_0}{k!}, \quad k = 1, 2, \dots, n.$$

$$P_{ОТК} = \frac{1}{1 + \sum_{k=1}^n \left(\frac{\lambda}{\mu}\right)^k \cdot \frac{1}{k!}}$$

Обозначим  $\frac{\lambda}{\mu} = \rho$  - коэффициент использования канала (среднее число заявок, приходящих в СМО за среднее время обслуживания). Тогда

$$\left. \begin{aligned} P_k &= \rho^k \cdot \frac{P_0}{k!} \\ P_{ОТК} &= \left[ 1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} \right]^{-1} \end{aligned} \right\} \quad (11.13)$$

(11.13) называется *формулой Эрланга*.

Зная  $P_k, \forall k = 0, \dots, n$ , можно найти все характеристики СМО.

1. Вероятность отказа. Заявка получает отказ, когда все каналы заняты

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

2. Вероятность того, что заявка будет принята – это вероятность того, что в системе есть хотя бы один свободный канал, поэтому относительная пропускная способность равна

$$q = 1 - P_{\text{отк}} = 1 - \frac{\rho^n}{n!} P_0.$$

3. Абсолютная пропускная способность

$$Q = \lambda q = \lambda \cdot \left(1 - \frac{\rho^n}{n!} P_0\right).$$

4. Среднее число занятых каналов  $\bar{K}$  (это в данном случае и среднее число заявок в системе  $\bar{z}$ ). Эту величину можно получить, разделив количество заявок, обслуживаемых в единицу времени всей системой на количество заявок, обслуживаемых в единицу времени одним каналом, т.е.

$$\bar{K} = \frac{Q}{\mu}$$

$$\bar{K} = \rho \cdot \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} P_0\right)$$

## 10) 11.11. Многоканальная СМО с ожиданием

В СМО  $n$  - каналов. На СМО поступает поток с интенсивностью  $\lambda$ . Интенсивность обслуживания одного канала  $\mu$ , число мест в очереди  $m$  (т.е. если заявка пришла, а в очереди уже  $m$  заявок, то она уходит из СМО). Перечислим все состояния СМО.

$$\left. \begin{array}{l} S_0 - \text{все каналы свободны;} \\ S_1 - \text{занят один канал;} \\ \dots \\ S_n - \text{заняты все } n \text{ каналов.} \end{array} \right\} \text{очереди нет}$$

$S_{n+1}$  – заняты  $n$  каналов; 1 заявка в очереди

...

$S_{n+r}$  – заняты  $n$  каналов;  $r$  заявок в очереди

...

$S_{n+m}$  – заняты  $n$  каналов;  $m$  заявок в очереди



Граф состояний будет иметь вид:

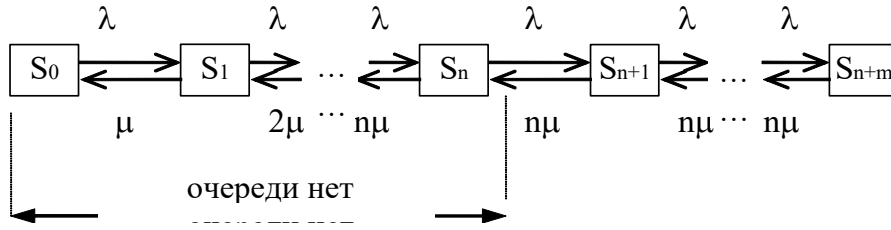


Рис. 89

Граф имеет вид схемы гибели и размножения. Напишем выражения для предельных вероятностей состояний, обозначив  $\frac{\lambda}{\mu} = \rho$ , используя формулы для схемы гибели и размножения (11.7), (11.8)

$$\begin{cases} P_1 = \frac{\rho}{1!} P_0; & P_2 = \frac{\rho^2}{2!} P_0; & \dots & P_n = \frac{\rho^n}{n!} P_0; \\ P_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} P_0; & P_{n+2} = \frac{\rho^{n+2}}{n^2 \cdot n!} P_0; & \dots & P_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} P_0; \end{cases} \quad (11.14)$$

$$P_{OTK} = \left[ \left( 1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} \right) + \underbrace{\frac{\rho^{n+1}}{n \cdot n!} + \frac{\rho^{n+2}}{n^2 \cdot n!} + \dots + \frac{\rho^{n+m}}{n^m \cdot n!}}_{\text{ααι ι άο δέ÷άηέαγ ι δι άδάνηεγ}} \right]^{-1}$$

Используем формулу суммы членов геометрической прогрессии со знаменателем  $q$ , получаем

$$S = \frac{a_1(1-q^m)}{1-q},$$

$$P_{OTK} = \left[ 1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} \cdot \frac{\frac{\rho}{n} - \left( \frac{\rho}{n} \right)^{m+1}}{1 - \frac{\rho}{n}} \right]^{-1} \quad (11.15)$$

Все вероятности состояний найдены. Найдем некоторые характеристики такой СМО:

1. Отказ будет тогда, когда заняты все каналы и нет места в очереди

$$P_{OTK} = P_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} P_0.$$

2. Относительная пропускная способность  $q$ , дополняет вероятность отказа до единицы

$$q = 1 - P_{OTK} = 1 - \frac{\rho^{n+m}}{n^m \cdot n!} P_0.$$

3. Абсолютная пропускная способность будет

$$Q = \lambda \cdot q = \lambda \cdot \left( 1 - \frac{\rho^{n+m}}{n^m \cdot n!} P_0 \right). \quad (11.16)$$

4. Найдем *среднее число заявок, связанных с системой*. Сначала найдем *среднее число занятых каналов*. Для СМО с отказами оно совпадало со средним числом заявок, находящихся в системе. Для нашего случая эти величины отличаются на величину заявок, находящихся в очереди.

Пусть  $\bar{Z}$  - число заявок, связанных с системой

$\bar{K}$  - среднее число занятых каналов

$$\bar{K} = \frac{Q}{\mu}$$

$$\bar{K} = \rho \cdot \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} P_0\right)$$

*Среднее число заявок в очереди* получим как математическое ожидание:

$$\begin{aligned} \bar{r} &= \sum_{r=1}^m r P_{n+r} = \\ &= \frac{\rho^{n+1} \cdot P_0}{n \cdot n!} \left[ 1 + 2 \frac{\rho}{n} + 3 \left( \frac{\rho}{n} \right)^2 + \dots + m \left( \frac{\rho}{n} \right)^{m-1} \right] \end{aligned}$$

$$\frac{\rho}{n} = \chi$$

$$\bar{r} = \frac{\rho^{n+1}}{n \cdot n!} P_0 \left[ 1 + 2\chi + 3\chi^2 + \dots + m\chi^{m-1} \right]$$

$$\left[ 1 + 2\chi + 3\chi^2 + \dots + m\chi^{m-1} \right] = \frac{d}{d\chi} (\chi + \chi^2 + \chi^3 + \dots + \chi^m) =$$

$$= \frac{d}{d\chi} \left( \frac{\chi - \chi^{m+1}}{1 - \chi} \right) = \frac{1 - \chi^m (m+1 - m\chi)}{(1 - \chi)^2},$$

$$\bar{r} = \frac{\rho^{n+1}}{n \cdot n!} P_0 \frac{1 - (m+1)\chi^m + m\chi^{m+1}}{(1 - \chi)^2}.$$

4. Очевидно, *среднее число заявок в системе* будет  $\bar{Z} = \bar{K} + \bar{r}$ .

5. Среднее время ожидания заявки в очереди. Здесь рассмотрим ряд гипотез:

- Заявка пришла, и очереди нет, есть свободный канал,  $t_{oc} = 0$ .
- Заявка пришла, и точно все  $n$  каналов заняты, очереди нет. Заявке придется ждать время  $\frac{1}{n\mu}$ , т.к. поток освобождений  $n$  каналов имеет интенсивность  $n\mu$ .
- Если заявка застанет все  $n$  - каналов занятыми и одну заявку в очереди, то ей придется ждать

$$\frac{2}{n\mu} = \frac{1}{n\mu} + \frac{1}{n\mu} \text{ и т.д.}$$

Если  $r$  заявок в очереди, то  $\frac{r}{n\mu}$ .

- Если в очереди  $m$  заявок, то она вообще не будет ждать.

$$\begin{aligned}\bar{t}_{i\alpha} &= \frac{1}{n\mu} P_n + \frac{2}{n\mu} P_{n+1} + \dots + \frac{m}{n\mu} P_{n+m-1} = \\ &= \frac{\rho^n P_0}{n \cdot n! \mu} \left[ 1 + \frac{2\rho}{n} + \frac{3\rho^2}{n^2} + \dots + \frac{m\rho^{m-1}}{n^{m-1}} \right]\end{aligned}$$

$$\bar{t}_{i\alpha} = \frac{\bar{r}}{\lambda}$$

$$\bar{t}_{i\alpha} = \frac{\rho^n P_0}{n \cdot n! \mu} \cdot \frac{1 - (m+1)\chi^m + m\chi^{m+1}}{(1-\chi)^2}$$

6. Среднее время пребывания заявки в системе

$$\bar{t}_{сист} = \bar{t}_{ож} + \frac{\text{относительная пропускная способность}}{\text{интенсивность обслуживания}}$$

$$\bar{t}_{сист} = \bar{t}_{ож} + \frac{q}{\mu}; \quad \left( \frac{1}{\mu} - \text{среднее время обслуживания} \right).$$

Действительно, т.к. заявки обрабатываются параллельно, то если одна заявка пришла, она обрабатывается за время  $\frac{1}{\mu}$ , но среди всех только  $q$  - часть обрабатывается действительно, поэтому  $\frac{q}{\mu}$ .

Таким образом, мы получили формулы для многоканальной СМО с ожиданием:

$$1. P_{отк} = \frac{\rho^{n+m}}{n^m \cdot n!} P_0$$

$$2. q = 1 - P_{отк}$$

$$3. Q = \lambda \cdot \left( 1 - \frac{\rho^{n+m}}{n^m \cdot n!} P_0 \right)$$

$$4. Z = \rho \cdot \left( 1 - \frac{\rho^{n+m}}{n^m \cdot n!} P_0 \right)$$

$$5. \bar{r} = \frac{\rho^{n+1}}{n \cdot n!} P_0 \cdot \frac{1 - (m+1)\chi^m + m\chi^{m+1}}{(1-\chi)^2}$$

$$6. \bar{K} = \bar{Z} + \bar{r}$$

$$7. \bar{t}_{ож} = \frac{\rho^n}{n \cdot \mu \cdot n!} P_0 \cdot \frac{1 - (m+1)\chi^m + m\chi^{m+1}}{(1-\chi)^2} = \frac{\bar{r}}{\lambda}$$

$$8. t_{сист} = \bar{t}_{ож} + \frac{q}{\mu};$$

$$P_0 = \left[ 1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} \cdot \frac{\left( \frac{\rho}{n} \right) - \left( \frac{\rho}{n} \right)^{m+1}}{1 - \left( \frac{\rho}{n} \right)} \right]^{-1} \quad (11.17)$$

## 11) 11.12. Одноканальная СМО с ожиданием

Очевидно, одноканальная СМО с ожиданием - есть частный случай многоканальной СМО с ожиданием, только что рассмотренной. Граф состояний

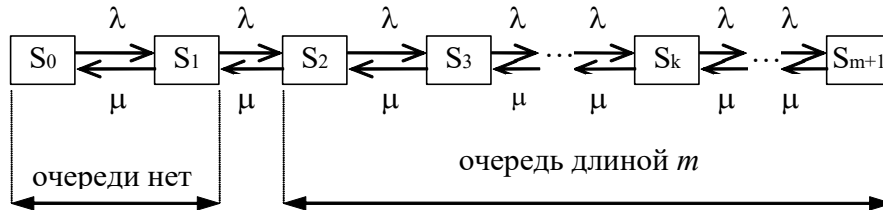


Рис. 90

Заметим, что для одноканальной системы, используя (11.14) и (11.15) получим

$$n=1: \quad P_k = \rho^k P_0; \quad P_0 = \left[ 1 + \rho + \rho^2 + \dots + \rho^{m+1} \right]^{-1} = \frac{1 - \rho}{1 - \rho^{m+2}}$$

Число мест в очереди ограничено  $m$ . Тогда, если подставить во все формулы, полученные в предыдущем параграфе  $n=1$ , получим:

$$\left. \begin{array}{l} P_1 = \rho \cdot P_0 \\ P_2 = \rho^2 \cdot P_0 \\ \dots \quad \dots \quad \dots \\ P_k = \rho^k \cdot P_0 \\ \dots \quad \dots \quad \dots \\ P_{m+1} = \rho^{m+1} \cdot P_0 \end{array} \right| \begin{array}{l} P_0 = \left[ 1 + \rho + \rho^2 + \dots + \rho^{m+1} \right]^{-1} \\ P_0 = \frac{1 - \rho}{1 - \rho^{m+2}} \end{array} \quad (11.18)$$

Получим характеристики такой СМО

$$1. \quad P_{отк} = P_{m+1} = \frac{\rho^{m+1}(1 - \rho)}{1 - \rho^{m+2}} \quad (11.19)$$

2. Относительная пропускная способность

$$q = 1 - P_{отк} = 1 - \frac{\rho^{m+1}(1 - \rho)}{1 - \rho^{m+2}} \quad (11.20)$$

3. Абсолютная пропускная способность

$$Q = \lambda \cdot q$$

4. Среднее число заявок в очереди. (Из формулы для многоканальной СМО при  $n=1$ )

$$\text{т.к. } \frac{\rho}{n} = \chi, \quad \text{при } n=1 \quad \chi = \rho$$

$$\bar{r} = \frac{\rho^2(1-\rho)}{1-\rho^{m+2}} \cdot \frac{1-\rho^m(m+1-m\rho)}{(1-\rho)^2} = \frac{\rho^2[1-\rho^m(m+1-m\rho)]}{(1-\rho^{m+2})(1-\rho)} \quad (11.21)$$

5. Среднее число заявок под обслуживанием, найдем из предыдущих формул при  $n=1$

$$\bar{K} = \rho \cdot \left(1 - \frac{\rho^{n+m}}{n^m n!} P_0\right) = \rho \cdot \left(1 - \frac{\rho^{m+1} \cdot (1-\rho)}{1-\rho^{m+2}}\right) = \frac{\rho - \rho^{m+2}}{1-\rho^{m+2}} \quad (11.22)$$

6. Среднее число заявок, связанных с СМО

$$\bar{Z} = \bar{r} + \bar{K}$$

7. Среднее время ожидания заявки в очереди

Для многоканальной СМО

$$\bar{t}_{ож} = \frac{\rho^n P_0}{n\mu n!} \cdot \frac{1 - (m+1)\chi^m + m\chi^{m+1}}{(1-\chi)^2}$$

При  $n=1$ :  $\chi = \rho$

$$\begin{aligned} \bar{t}_{ож} &= \frac{\rho(1-\rho)}{\mu(1-\rho^{m+2})!} \cdot \frac{1 - (m+1)\rho^m + m\rho^{m+1}}{(1-\rho)^2} = \\ &= \frac{\rho}{\mu(1-\rho^{m+2})!} \cdot \frac{1 - \rho^m(m+1-m\rho)}{(1-\rho)} \end{aligned} \quad (11.23)$$

Сравнив это выражение с выражением для  $\bar{r}$ , видим, что

$$\bar{t}_{ож} = \frac{1}{\rho \cdot \mu} \bar{r} = \frac{\bar{r}}{\lambda}, \quad (11.24)$$

т.е. среднее время ожидания равно среднему числу заявок в очереди, деленному на интенсивность потока заявок

$$\bar{t}_{сист} = \bar{t}_{ож} + \frac{q}{\mu} = \frac{\bar{r}}{\lambda} + \frac{q}{\mu}$$

## 12) 11.13.СМО с ожиданием и бесконечной очередью

Мы рассмотрели СМО с ожиданием и с очередью длины  $m$ . Снимем это ограничение и устремим  $m \rightarrow \infty$ . Граф состояний будет бесконечным. Получим

вероятности состояний одноканальной СМО с бесконечной очередью. Для этого в выражениях (11.18) - (11.23), положим  $t \rightarrow \infty$ . Видим, что в выражении для  $P_0$  сумма будет конечной, только если  $\rho < 1$ . Т.о. при  $\rho \geq 1$  не существует предельного установившегося режима, и при  $t \rightarrow \infty$  очередь стремится к бесконечности.

Пусть  $\rho = \frac{\lambda}{\mu} < 1$  и предельный режим существует, тогда (используя сумму убывающей геометрической прогрессии) получим

$$\left. \begin{aligned} P_0 &= 1 - \rho \\ P_1 &= \rho \cdot (1 - \rho) \\ \dots &\dots \dots \\ P_k &= \rho^k \cdot (1 - \rho) \\ \text{Каждая заявка будет обслужена, поэтому} \\ 1. \quad P_{отк} &= 0 \\ 2. \quad q &= 1 \\ 3. \quad Q &= \lambda \cdot q = \lambda \\ 4. \quad \bar{r}_\infty &= \frac{\rho^2}{1 - \rho} \\ 5. \quad \bar{K}_\infty &= \rho \\ 6. \quad \bar{Z}_\infty &= \bar{r}_\infty + \rho = \frac{\rho^2}{1 - \rho} + \rho = \rho \left( \frac{\rho}{1 - \rho} + 1 \right) = \frac{\rho}{1 - \rho} \\ 7. \quad \bar{t}_{ож} &= \frac{1}{\mu} \cdot \frac{\rho}{1 - \rho} \\ 8. \quad \bar{t}_{сист} &= \frac{1}{\mu} \cdot \frac{\rho}{1 - \rho} + \frac{1}{\mu} = \frac{1}{\mu} \cdot \frac{1}{1 - \rho} \end{aligned} \right\} (11.25)$$

### 13) 11.14. Замкнутые системы массового обслуживания

Мы рассмотрели СМО, когда заявки поступали извне. Это *разомкнутые (открытые)* СМО. Иногда потоки заявок зависят от состояния самой СМО. Рассмотрим пример такой *замкнутой СМО*.

Бригада из  $m$  рабочих обслуживает  $n$  станков ( $m < n$ ). При выходе из строя станка один из рабочих его обслуживает (если есть свободные рабочие). Если вышли из строя  $(m+1)$  станков, то один станок не может быть немедленно обслужен, он становится в очередь. Считаем что распределение времени обслуживания – пуассоновское. Рабочий в среднем тратит время на обслуживание

$$\bar{t}_{об} = \frac{1}{\mu},$$

где  $\mu$  - интенсивность потока обслуживаний.

Пусть  $\lambda$  - интенсивность потока неисправностей одного станка. Перечислим состояния:

$S_0$  – все станки работают, рабочие свободны

$S_1$  – один станок неисправен, один рабочий его ремонтирует

...

$S_m$  –  $m$  станков неисправны,  $m$  рабочих их ремонтируют.

...

$S_n$  –  $n$  станков неисправны,  $m$  станков ремонтируются, а  $n-m$  станков ожидают ремонта.

Граф состояний будет иметь вид.

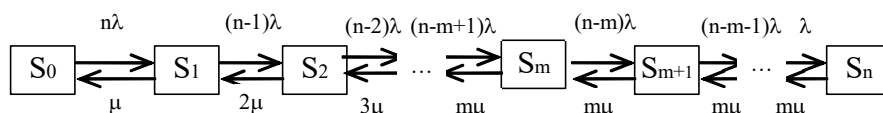


Рис. 91

Видим, что это схема гибели и размножения. Используя общее решение (11.7), (11.8), получим предельные вероятности состояний.

$$P_1 = \frac{n}{1} \rho \cdot P_0$$

$$P_2 = \frac{n(n-1)}{2!} \rho^2 \cdot P_0$$

$$P_3 = \frac{n(n-1)(n-2)}{3!} \rho^3 \cdot P_0$$

...

$$P_m = \frac{n(n-1)\dots(n-m+1)}{m!} \rho^m \cdot P_0$$

$$P_{m+1} = \frac{n(n-1)\dots(n-m)}{m \cdot m!} \rho^{m+1} \cdot P_0$$

$$P_{m+2} = \frac{n(n-1)\dots(n-m-1)}{m^2 \cdot m!} \rho^{m+2} \cdot P_0$$

...

$$P_n = \frac{n(n-1)\dots 1}{m^{n-m} \cdot m!} \rho^n \cdot P_0$$

$$P_0 = \left[ 1 + \frac{n}{1!} \rho + \frac{n(n-1)}{2!} \rho^2 + \dots + \frac{n(n-1)\dots(n-m+1)}{m!} \rho^m + \right. \\ \left. + \frac{n(n-1)\dots(n-m)}{m \cdot m!} \rho^{m+1} + \dots + \frac{n(n-1)\dots 1}{m^{n-m} \cdot m!} \rho^n \right]^{-1}$$

Вычислим среднее число занятых рабочих

$$\bar{K} = 0 \cdot P_0 + 1 \cdot P_1 + 2 \cdot P_2 + \dots + m \cdot (P_m + P_{m+1} + \dots + P_n)$$

$$\bar{K} = P_1 + 2 \cdot P_2 + \dots + (m-1) \cdot P_{m-1} + m \cdot (1 - P_0 - P_1 - \dots - P_{m-1})$$

*Среднее число неисправностей, обслуживаемое бригадой в единицу времени (абсолютная пропускная способность) получается как произведение среднего числа занятых рабочих на количество обслуживаемых заявок в единицу времени*

$$Q = \bar{K} \cdot \mu$$

*Среднее число неисправных станков, т.е. среднее число станков, связанных с процессом обслуживания -  $\varpi$  (аналогично параметру  $\bar{z}$ ). Вычислим  $\varpi$  непосредственно как математическое ожидание*

$$\varpi = 1 \cdot P_1 + 2 \cdot P_2 + \dots + m \cdot P_m + (m+1) \cdot P_{m+1} + \dots + n \cdot P_n$$

Но можно получить другую формулу, не требующую знания всех вероятностей.

Действительно: количество работающих станков  $(n - \varpi)$ . Эти работающие станки создают интенсивность потока неисправностей  $(n - \varpi)\lambda$ . В стационарном режиме эти неисправности устраняются рабочими, которые устраняют  $Q$  неисправностей в единицу времени. Значит

$$(n - \varpi)\lambda = Q = \bar{K} \cdot \mu$$

$$n - \varpi = \frac{\bar{K} \cdot \mu}{\lambda}$$

$$\varpi = n - \frac{\bar{K} \cdot \mu}{\lambda} = n - \frac{\bar{K}}{\rho}$$

*Среднее число станков, ожидающих в очереди*

$$\bar{r} = \varpi - \bar{K}$$

*Вероятность того, что рабочие не будут заняты*

$$P_{своб} = P_0$$

#### 14) 11.15. СМО с взаимопомощью между каналами

Иногда можно построить СМО, где незанятые каналы помогают занятому в обслуживании. Это может быть и в открытых, и в замкнутых системах массового обслуживания.

При анализе операций в подобных СМО необходимо учитывать следующее:

1. Насколько убыстряется обслуживание при взаимопомощи.
2. Какой регламент (алгоритм) взаимопомощи дает наилучшие результаты.



Если над обслуживанием заявки работает не один канал, а  $k$  каналов, то интенсивность потока обслуживания будет увеличиваться с ростом  $k$ . Обозначим эту функцию  $\mu(k)$ :

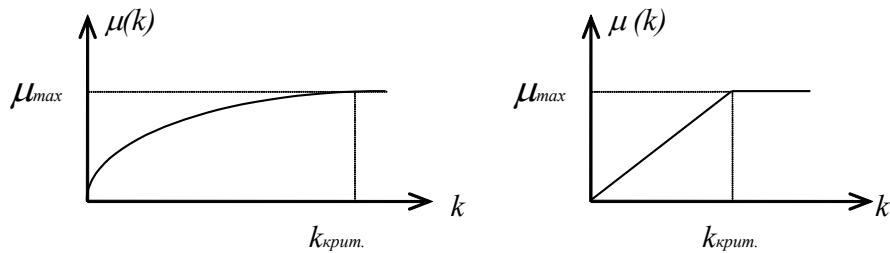


Рис. 92

Очевидно, что неограниченное увеличение числа помогающих каналов не всегда ведет к пропорциональному увеличению скорости обслуживания. При некотором  $k=k_{\text{крит}}$  дальнейшее увеличение числа занятых каналов бессмысленно.

Для анализа поэтому надо знать  $\mu(k)$ .

Пусть  $\mu(k)=ak \quad k \leq k_{\text{крит}};$

$\mu(k)=\mu_{\text{max}} \quad k > k_{\text{крит}};$

Если  $n \leq k_{\text{крит}}$ , то интенсивность обслуживания заявки несколькими каналами пропорционально числу каналов.

Обратимся к анализу алгоритмов взаимопомощи.

Самый простой алгоритм «все как один» на заявку. В этом случае все  $n$  каналов обслуживают появившуюся заявку до полного обслуживания, затем вторую и т.д. В этом случае СМО как бы становится одноканальной, только с более высокой интенсивностью обслуживания.

Выгодно ли это? Это зависит от  $\mu(k)$ , от типа СМО (с отказами, с очередью) и от критериальной величины.

Очевидно, для систем с отказами *среднее время пребывания заявки* в СМО уменьшается. Это хорошо, зато увеличивается *вероятность отказа*, т.к. все каналы заняты, если хотя бы одна заявка пришла.

Для системы с ожиданием также получаем одноканальную СМО с увеличенной интенсивностью обслуживаний. Здесь *среднее время пребывания заявки в системе* меньше, но *средняя длина очереди* и *среднее время ожидания в очереди* может получиться больше.

Таким образом, алгоритм “все как один” не очень хорош. Желательно сделать так, чтобы взаимопомощь между каналами не мешала принимать и обслуживать новые заявки, появившиеся за время, пока все каналы были заняты.

Назовем “распределенной взаимопомощью” следующий алгоритм:

В момент прихода первой заявки за обслуживание принимаются все  $n$  каналов, если приходит еще одна, то часть каналов переключается на нее, если еще одна, то часть на нее и т.д. до тех пор, пока не окажутся занятыми  $n$  каналов. Далее происходит отказ в СМО с отказами, или заявки становятся в очередь в СМО с очередями. Здесь нет простоя, если хотя бы одна заявка поступила.

Возникает вопрос: какую часть каналов надо перераспределять на другие, поступившие заявки. Это зависит от вида  $\mu(k)$ . Если  $\mu(k)$  линейна и  $k_{\text{крит}} > n$ , то все равно какую часть каналов выделить на обслуживание вновь поступившей заявки, лишь бы все каналы были заняты (суммарная

интенсивность обслуживаний все равно будет  $n\mu$ ). В общем случае для нахождения оптимального распределения каналов по заявкам нужно решать задачу

$$\left. \begin{aligned} \sum_{i=1}^l \mu(k_i) &\rightarrow \max \\ \sum_{i=1}^l k_i &= n \end{aligned} \right\} \quad (11.26)$$

Эта задача является частным случаем задачи распределения ресурсов (см.п.5.4) и, как известно, имеет теоретическое решение.

Если кривая  $\mu(k)$  неубывающая и выпуклая вверх, то нужно распределять каналы по заявкам как можно более равномерно, если выпуклая вниз, то наиболее неравномерно – все на одну заявку. (Если зависимость линейная, то все равно).

Рассмотрим работу  $n$  - канальной СМО при “равномерной” взаимопомощи между каналами.

#### 15) 11.15.1 Взаимопомощь в СМО с отказами.

$S_0$  - СМО свободна;

$S_1$  - одна заявка обслуживается всеми  $n$  каналами;

$S_2$  - две заявки обслуживаются всеми  $n$  каналами;

.....

$S_k$  -  $k$  заявок обслуживается всеми  $n$  каналами;

.....

$S_n$  -  $n$  заявок обслуживается всеми  $n$  каналами;

Граф состояний будет таким

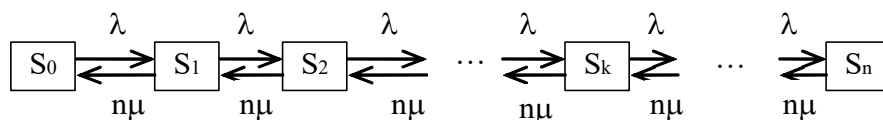


Рис. 93

Т.е. граф такой же как для одноканальной СМО с  $\mu^*=n\mu$  и ограниченной очередью, имеющей  $(n-1)$  мест (см.п.11.12). Поэтому можно подставить в полученные ранее для этой СМО формулы вместо  $\rho$

$$\rho^* = \frac{\lambda}{\mu^*} = \frac{\lambda}{n\mu}$$

и вместо  $m$  величину

$$m^*=n-1.$$

#### 16) 11.15.2 Взаимопомощь в СМО с очередью.

Рассмотрим СМО с очередью и максимальным числом мест в очереди  $m$ . Пусть между каналами равномерная взаимопомощь и  $\mu(k)=k\mu$ .

$S_0$  - СМО свободна;

$S_1$  - одна заявка обслуживается всеми  $n$  каналами;

$S_2$  - две заявки обслуживаются всеми  $n$  каналами;

.....

$S_k$  -  $k$  заявок обслуживается всеми  $n$  каналами, очереди нет;

.....

$S_n$  -  $n$  заявок обслуживается всеми  $n$  каналами, очереди нет;

.....

$S_{n+1}$  -  $n$  заявок обслуживается всеми  $n$  каналами, одна заявка стоит в очереди;

.....

$S_{n+m}$  -  $n$  заявок обслуживается всеми  $n$  каналами, в очереди  $m$  заявок;

Граф состояний будет таким

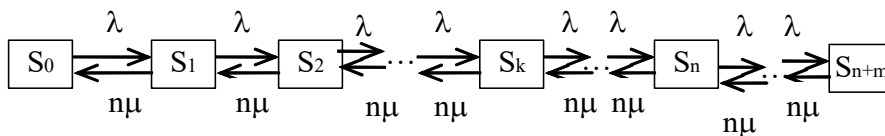


Рис. 94

Мы получили граф того же вида, что и в предыдущем случае, но с увеличенным на  $m$  числом состояний, т.е. здесь также можно воспользоваться формулами для одноканальной СМО с ограниченной очередью с  $\mu^* = n\mu$  и числом мест в очереди  $m^* = n + m - 1$ . Если записать все характеристики, то увидим, что для равномерной взаимопомощи все характеристики изменяются в желательном для нас направлении. Так как интенсивность и число мест в очереди увеличивается, то  $P_{отк}$  уменьшается, а  $q$  и  $Q$  увеличивается.

Однако организация такой дисциплины взаимопомощи осуществима не для всех СМО.

## 17) 11.16. СМО с ошибками

На практике часто удобно использовать модель СМО, в которой заявки обслуживаются не с полной достоверностью, а с вероятностью  $p \neq 1$ , т.е. некоторые заявки в действительности не обслуживаются (брак в работе СМО, который нельзя принципиально устранить).

Например: СМО – операция по отражению налетов противника, справочное бюро, телефонная станция и т.п.

В разомкнутой СМО с ошибками появление ошибки на обслуживание практически не сказывается, т.к. число источников заявок очень велико и интенсивность потока заявок из-за ошибок практически не меняется. Поэтому, чтобы учесть ошибки в разомкнутой системе, нужно умножить на  $p$  относительную пропускную способность  $q$ , где  $p$  - вероятность безошибочного обслуживания. Соответственно  $Q^* = Qp$ . Другие характеристики, такие как время ожидания в очереди, число заявок в очереди, очевидно, не меняются.

Для замкнутой системы, заявка, обслуженная с ошибкой, вновь становится в очередь на обслуживание и увеличивает время обслуживания всех заявок.

Пример.

Рассмотрим замкнутую СМО с ошибками. Пусть один рабочий обслуживает  $n$  станков. Интенсивность потока неисправностей одного станка  $\lambda$ , среднее время наладки  $\bar{t}_{ос} = \frac{1}{\mu}$ . С вероятностью  $p$  обслуживание удачно. С вероятностью  $(1-p)$

обслуживание с браком. Требуется определить предельные вероятности состояний.

Здесь наличие ошибки приводит к тому, что  $\mu^* = p\mu$ , т.к. число успешных обслуживаний за единицу времени будет именно таким.

Получим граф состояний рис. 95

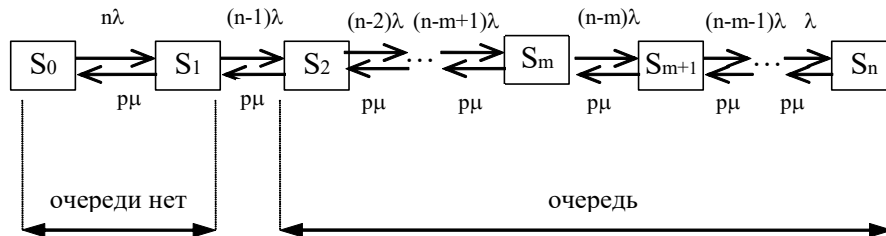


Рис. 95

Мы получили граф состояний как в одноканальной замкнутой СМО. По полученным ранее формулам можно получить предельные вероятности.

Важной является модель СМО, когда *вероятность обслуживания* зависит от длины очереди. Пусть в этом же примере при отсутствии очереди интенсивность потока обслуживания

$$\mu(0) = \frac{1}{\bar{t}_{об}}$$

При наличии очереди  $r$  интенсивность  $\mu(r)$  обычно больше, чем  $\mu(0)$  ( $\mu(r) > \mu(0)$ ). Одновременно изменяется  $p(r) > p(0)$ , тогда интенсивность верного обслуживания будет

$$\mu^*(r) = \mu(r)p(r), \quad (r=0, \dots, n-1).$$

Граф состояний такой

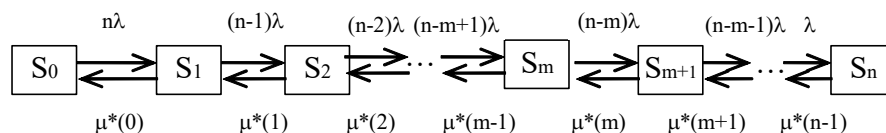


Рис. 96

Получаем вероятности состояний как для схемы гибели и размножения:

$$P_1 = \frac{n\lambda}{\mu^*(0)} P_0$$

$$P_2 = \frac{n(n-1)\lambda^2}{\mu^*(0) \cdot \mu^*(1)} P_0$$

... ..

$$P_n = \frac{n(n-1)\dots 1 \cdot \lambda^n}{\mu^*(0)\dots \mu^*(n-1)} P_0$$

$$P_0 = \left[ 1 + \frac{n\lambda}{\mu^*(0)} + \frac{n(n-1)\lambda^2}{\mu^*(0) \cdot \mu^*(1)} + \dots + \frac{n(n-1)\dots 1 \cdot \lambda^n}{\mu^*(0)\dots \mu^*(n-1)} \right]^{-1}$$

### 13.9. СМО с ограниченным временем ожидания

Мы рассматриваем модели СМО, где система с ожиданием при заданном числе мест в очереди  $m$ . Если заявка попала в очередь, то она рано или поздно обслужится – «терпеливая заявка».

На практике заявка не ждет больше какого-то времени – «нетерпеливая заявка».

Пример:  $n$  каналов, число мест в очереди не ограничено. Время пребывания заявки в очереди ограничено случайным временем  $T_{оч}$ .

$T_{оч}$  – нахождение в очереди. Распределение пуассоновское.

$$\nu = \frac{1}{t_{оч}}$$

На каждую заявку действует поток ухода с интенсивностью:

Рассчитаем характеристики СМО:

$S_1$  – занят один канал

$S_n$  – все каналы

$S_{n+1}$  – 1 в очереди

$S_{n+r}$  –  $r$  в очереди

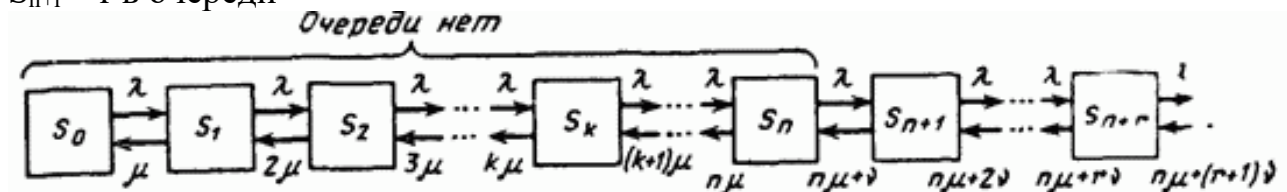


Схема гибели и размножения.

$$\rho = \lambda/\mu \quad \beta = \nu/\mu \quad \rho = \lambda/\mu, \quad \beta = \nu/\mu,$$

Особенности:

Если длина очереди неограниченна, а заявки терпеливы, то стационарный предельный режим существует в случае  $\rho < n$ . При  $\rho \geq n$  ряд неограничен (расходится), очередь увеличивается. А с нетерпеливыми заявками система будет работать даже при  $\rho \geq n$ . Для СМО с нетерпеливыми заявками такой показатель, как вероятность отказа не имеет смысла.

1. **Относительная пропускная способность**  $q = Q/\lambda = 1 - \nu\bar{r}/\lambda$

(A это Q, K – z)

2. Среднее число занятых каналов

(A это Q, K – z)

3. Среднее число заявок в очереди

(A это Q, K – z)

$$q = \frac{A}{\lambda} = \frac{\lambda - v\bar{r}}{\lambda} = 1 - \frac{v}{\lambda} \bar{r}.$$

$$\bar{k} = Q/\mu = \lambda/\mu - v\bar{r}/\lambda = \rho - \beta\bar{r}$$

$$\bar{z} = \frac{A}{\mu} = \frac{\lambda - v\bar{r}}{\mu} = \rho - \beta\bar{r}.$$

$$\bar{r} = \rho/\beta - \bar{k}/\beta$$

$$\bar{r} = \frac{\rho}{\beta} - \frac{\bar{z}}{\beta},$$

### 13.10. СМО с непуассоновскими потоками (11.17)

#### 18) 11.17. СМО с произвольным распределением времени обслуживания. Формула Хинчина-Поллачека

Пуассоновский поток является важнейшей моделью для анализа большинства СМО. Роль его в теории массового обслуживания аналогична роли нормального закона распределения в теории вероятностей, т.к. сумма большого числа потоков даже отличных от пуассоновских дает пуассоновский поток.

Анализ СМО, когда потоки отличаются от пуассоновских сильно усложняется.

Остановимся на одном важнейшем случае, когда анализ позволяет получить характеристики СМО при произвольном вероятностном распределении потоков в СМО.

Пусть в одноканальной СМО с бесконечным числом мест в очереди ( $m=\infty$ ) поток заявок пуассоновский, а поток обслуживания произвольный, но известно среднее значение времени обслуживания и его постоянная дисперсия

$$\bar{t}_{об} = \frac{1}{\mu}, \quad \sigma_{t_{об}}^2 = \int_0^{\infty} \left(t - \frac{1}{\mu}\right)^2 \cdot g(t) dt$$

где  $g(t)$  - произвольная плотность распределения длительностей обслуживания.

Тогда, для любого вида  $g(t)$  средняя длина очереди определяется выражением

$$\bar{r} = \frac{\rho^2}{2(1-\rho)} \left\{ 1 + \left( \frac{\sigma_{t_{об}}}{\bar{t}_{об}} \right)^2 \right\} = \frac{\rho^2(1 + \mu^2 \sigma_{t_{об}}^2)}{2(1-\rho)} = \frac{\rho^2 + \lambda^2 \sigma_{t_{об}}^2}{2(1-\rho)} \quad (11.27)$$

Это, так называемая, формула Хинчина-Поллачека.

Среднее число заявок в системе, очевидно, будет

$$\bar{K} = \bar{Z} + \bar{r}$$

$$\bar{K} = \rho + \bar{r} = \rho + \frac{\rho^2 + \lambda^2 \sigma_{t_{об}}^2}{2(1-\rho)}$$

Для дисциплины – “первый пришел – первый обслуживается” среднее время пребывания заявки в системе будет [8]

$$\bar{t}_{сист} = \bar{t}_{ож} + \bar{t}_{об} = \frac{1}{\lambda} \left[ \rho + \frac{\rho^2 + \lambda^2 \sigma_{t_{об}}^2}{2(1-\rho)} \right]$$

*Относительное среднее время простоя системы*

$$\bar{t}_{простоя} = P_0 = 1 - \rho$$

Формула Хинчина-Поллачека верна для любой дисциплины управления очередью например “первый пришел - первый ушел”, “последний пришел - первым обслуживается”, лишь бы выбор очередной заявки для обслуживания не зависел от времени обслуживания.

Рассмотрим два важных случая:

1. Время обслуживания постоянно (самый легкий для обслуживания случай). В этом случае

$$\begin{aligned} \sigma_{t_s} &= 0, \\ \bar{r} &= \frac{\rho^2}{2(1-\rho)} \\ \bar{K} &= \rho + \frac{\rho^2}{2(1-\rho)}; \end{aligned} \tag{11.28}$$

2. Время обслуживания случайное и распределение имеет экспоненциальный характер, т.е. поток обслуживания - пуассоновский (самый тяжелый для обслуживания случай). На основании (11.6)

$$\sigma_{\bar{t}_{об}} = \bar{t}_{об} = \frac{1}{\mu},$$

Поставляя в формулу Хинчина-Поллачека получаем

$$\begin{aligned} \bar{r} &= \frac{\rho^2}{2(1-\rho)} (1+1) = \frac{\rho^2}{1-\rho} \\ \bar{K} &= \rho + \frac{\rho^2}{1-\rho} = \frac{\rho}{1-\rho}; \end{aligned} \tag{11.29}$$

Т.о. формулы для пуассоновского потока обслуживания (см.п.11.13) можно получить как частный случай формулы Хинчина-Поллачека.

Сравнивая (11.18) и (11.19) замечаем, что средняя длина очереди для пуассоновского потока в 2 раза больше, чем для постоянного времени обслуживания.

Поэтому, если мы рассчитываем на пуассоновский поток обслуживания, а на самом деле поток отличается от пуассоновского, то средняя длина очереди будет

несколько завышена (максимум в два раза), но это страхует нас от ошибки в другую сторону.

Сказанное приводит нас к выводу о том, что, рассчитывая характеристики СМО в предположении пуассоновских потоков, мы получаем предельные (наихудшие) характеристики. Это значит, что они гарантируют нам, что в реальных СМО характеристики будут не хуже.

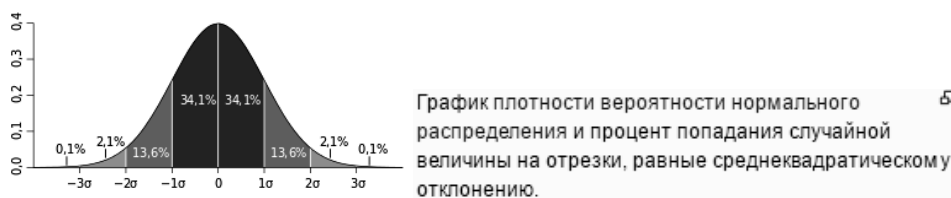
В этом заключается роль пуассоновской модели потоков для СМО.

Анализ более сложных моделей СМО, применение метода имитационного моделирования для анализа, а также практическое применения моделей СМО смотри в [1,8,10].

## 11. Типовые распределения случайных величин (СВ), используемые в моделях СМО

Мы рассмотрели получение вероятностей всех состояний для системы с простейшими потоками. На практике встречаются системы с другими распределениями потоков, при этом, для некоторых распределений ставится задача свести к эквивалентному с простейшими потоками. Системе с простейшими потоками удастся получить характеристики в аналитическом виде. Характеристики СМО зависят от свойств потоков, которые определяются видами распределения и параметрами (среднее, дисперсия и др.). Именно для СМО основным параметром распределения потоков является безразмерная величина, которая называется **коэффициентом вариации** случайной величины  $X$ , обозначаемой  $V(x)$ .

**Коэффициент вариации** определяется как **отношение СКО к мат ожиданию**  $v(x)=\sigma(x)/M(x)$ .



Пример с студенческим буфетом означает, что для простейшего потока события происходят неравномерно (то густо, то пусто).

### I. Дискретные случайные величины (ДСВ)

#### 1) Пуассоновское распределение

Пуассоновской называется целочисленная случайная величина, множество возможных значений которой:  $X = \{0, 1, 2, \dots, n, \dots\} = \{x_k = k, k = 0, 1, 2, \dots\}$ , а вероятности, с

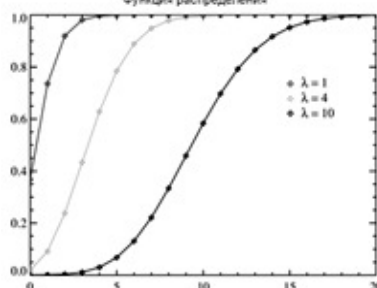
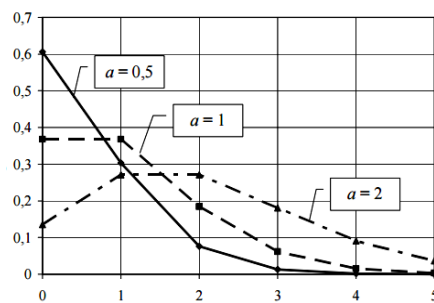
которыми значения принимаются, задаются формулой:  $p_k = P(X = k) = \frac{a^k}{k!} e^{-a}, k = 0, 1, 2, \dots$ .

Число  $a > 0$  называется параметром пуассоновской случайной величины.



$X$	0	1	...	$n$	...
$P$	$e^{-a}$	$ae^{-a}$	...	$\frac{a^n}{n!}e^{-a}$	...

$$MX = a, DX = a$$



## 2) Геометрическое распределение

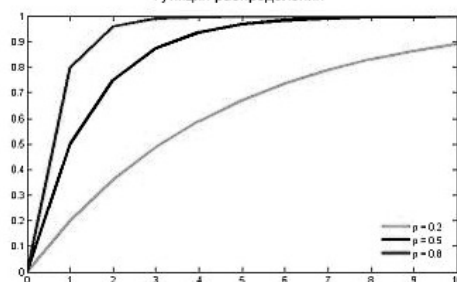
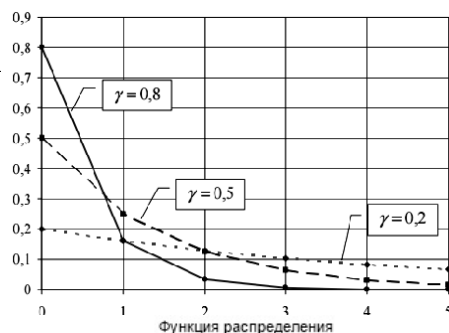
Геометрической называется дискретная случайная величина  $X$ , представляющая собой число испытаний, проводимых по схеме Бернулли, до появления первого успеха с вероятностью успеха в одном испытании равной  $p$ . Геометрическая случайная величина имеет счетное множество возможных значений:

$X = \{1, 2, \dots, n, \dots\} = \{x_k = k, k = 1, 2, \dots\}$ . Вероятности значений определяются по формуле:

$$p_k = P(X = k) = q^{k-1}p, k = 1, 2, \dots$$

$X$	1	2	...	$n$	...
$P$	$p$	$qp$	...	$q^{n-1}p$	...

$$MX = \frac{1}{p}, DX = \frac{q}{p^2}$$



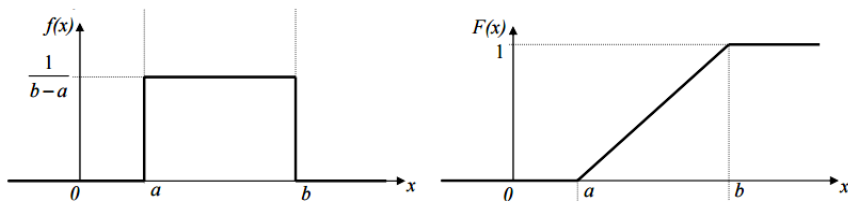
## II. Непрерывные случайные величины

Поскольку в СМО во всех моделях случайной величиной (СВ) является время, то присутствует и непрерывная СВ.

### 1.Равномерное распределение

Непрерывная СВ  $X$  имеет равномерный закон распределения (равномерное распределение) на отрезке  $[a, b]$ , если множество ее возможных значений  $X = [a, b]$ , а плотность вероятностей постоянна на этом отрезке. Интегрируя плотность вероятностей получим функцию распределения:

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & x \notin [a, b], \end{cases} \quad F_X(x) \equiv \mathbb{P}(X \leq x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} \quad MX = \frac{a+b}{2}, \quad DX = \frac{(b-a)^2}{12}.$$

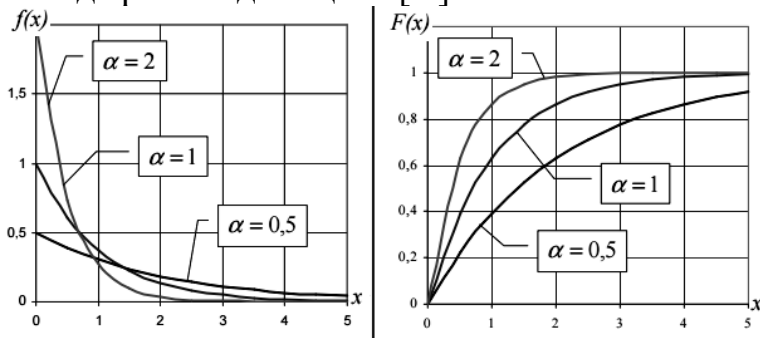


## 2. Экспоненциальное распределение (СВ на $[0, +\infty)$ )

Непрерывная СВ  $X$  имеет показательный закон распределения (показательное, экспоненциальное распределение), если множество ее возможных значений  $X = [0, +\infty)$ , а плотность вероятностей:

$$f(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases} \quad F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases} \quad MX = \frac{1}{\lambda}, \quad DX = \frac{1}{\lambda^2}$$

$\lambda$  – параметр распределения. Замечательной особенностью экспоненциального распределения является то, что его коэффициент вариации не зависит от параметра  $\alpha$  и всегда равен единице.  $V[X]=1$



Это распределение тесно связано со свойствами Марковских процессов. Переходы между состояниями не зависят от предыстории.

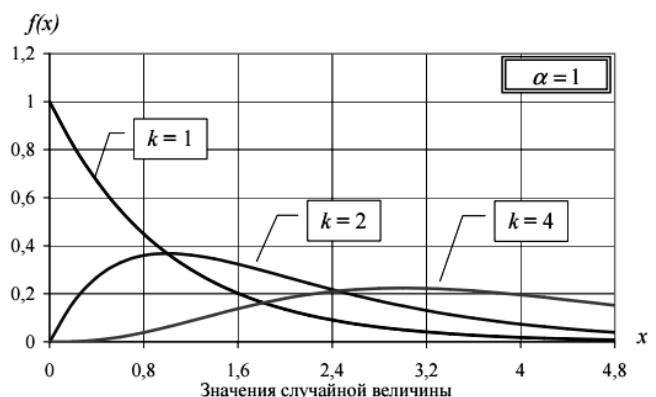
Возможность получения сравнительно простых аналитических результатов при использовании предположения об экспоненциальном характере случайных процессов обусловила появление рассматриваемых ниже специфических законов распределений, представляющих собой композиции экспоненциальных распределений и позволяющих упростить решение многих задач, связанных с исследованием моделей массового обслуживания. Это позволит упростить моделирование и получение характеристик систем.

## 3. Распределение Эрланга $k$ -ого порядка

НСВ принадлежит  $(0, \infty)$  и представляет сумму  $k$  независимых СВ, распределённых по одному и тому же закону с параметром  $\lambda$ . Другими словами, поток Эрланга получается, если убрать из Пуассоновского каждое  $k$ -ое событие. (Вместо альфа писать  $\lambda$ )

$$F_k(x) = 1 - e^{-\alpha x} \sum_{i=0}^{k-1} \frac{(\alpha x)^i}{i!};$$

$$f_k(x) = \frac{\alpha(\alpha x)^{k-1}}{(k-1)!} e^{-\alpha x};$$



При  $k=1$  распределение Эрланга вырождается в экспоненциальное, а при  $k \rightarrow \infty$  – приближается к нормальному распределению. Распределение Эрланга двухпараметрическое. Коэффициент вариации распределения Эрланга зависит от параметра  $k$  и принимает значения меньше или равное единице:

$$\nu_{\text{Эк}}[X] = \frac{1}{\sqrt{k}} \leq 1 \quad (k=1,2,\dots).$$

#### 4. Нормированное распределение Эрланга

Это распределение суммы  $k$  независимых случайных величин, каждая из которых распределена по экспоненциальному закону с параметром  $k\alpha$ , зависящим от  $k$ . Другими словами, суммируются  $k$  экспоненциально распределенных случайных величин, каждая из которых имеет математическое ожидание в  $k$  раз меньшее, чем исходное математическое ожидание реального распределения, что приводит к независимости математического ожидания нормированного распределения Эрланга от параметра  $k$ .

$$F_k(x) = 1 - e^{-k\alpha x} \sum_{i=0}^{k-1} \frac{(k\alpha x)^i}{i!};$$

$$f_k(x) = \frac{k\alpha (k\alpha x)^{k-1}}{(k-1)!} e^{-k\alpha x},$$

Коэффициент вариации нормированного распределения Эрланга так же, как и ненормированного, зависит от параметра  $k$  и принимает значения меньше или равное единице:

$$\nu_{\text{нЭк}}[X] = \frac{1}{\sqrt{k}} \leq 1 \quad (k=1,2,\dots).$$

#### 5. Гиперэкспоненциальное распределение

Когда некоторое реальное распределение непрерывной случайной величины, принимающей неотрицательные значения, имеет коэффициент вариации больше единицы, для его аппроксимации может использоваться гиперэкспоненциальное распределение. Гиперэкспоненциальное распределение некоторым образом связано с экспоненциальным и представляет собой аддитивную смесь разных экспоненциальных распределений. Процесс формирования случайных величин с этим распределением из экспоненциально распределенных случайных величин может быть представлен так:

Положим, что имеется  $n$  разных генераторов экспоненциально распределенных случайных величин с параметрами  $\alpha_1, \dots, \alpha_n$ , соответственно (мат. ожиданиями  $M_1 = 1/\alpha_1, \dots, M_n = 1/\alpha_n$ ), причем  $\alpha_i \neq \alpha_j$  для всех  $i \neq j$  ( $i, j = \text{от } 1 \text{ до } n$ ). Пусть в результате одного опыта с вероятностью  $q_i$  вырабатывается только одна случайная величина  $i$ -м генератором с параметром  $\alpha_i$  ( $i = \text{от } 1 \text{ до } n$ ), причем  $q_1 + \dots + q_n = 1$ .

Совокупность случайных величин, полученных в результате проведения множества таких опытов, будет распределена по гиперэкспоненциальному закону:



$$\left. \begin{aligned} F(x) &= \sum_{i=1}^n q_i (1 - e^{-\alpha_i x}) = 1 - \sum_{i=1}^n q_i e^{-\alpha_i x}; \\ f(x) &= \sum_{i=1}^n q_i \alpha_i e^{-\alpha_i x} \end{aligned} \right\}$$

В простейшем варианте случайные величины с гиперэксп распределением могут быть получены с использованием только двух экспоненциальных распределений:  $n=2$ . Тогда функция и плотность гиперэксп распределения будут иметь вид:

$$\left. \begin{aligned} F(x) &= qp(1 - e^{-\alpha_1 x}) + (1 - q)(1 - e^{-\alpha_2 x}); \\ f(x) &= q\alpha_1 e^{-\alpha_1 x} + (1 - q)\alpha_2 e^{-\alpha_2 x}. \end{aligned} \right\}$$

Плотность гиперэксп распределения по сравнению с экспоненциальным распределением характеризуется более резким спадом в области малых значений случайной величины, причем чем больше коэффициент вариации случайной величины, тем круче эта зависимость.

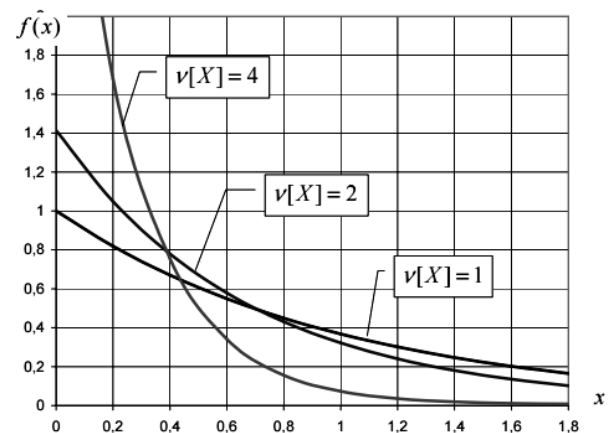
Можно показать, что вероятность появления маленьких значений случайной величины для гиперэксп распределения намного больше вероятности появления больших значений.

Определим вероятность того, что случайная величина примет значение меньше математического ожидания  $M$ .

Для этого рассчитаем значение функции распределения в точке  $x=M$ :  $P(X < M) = F(x = M) = F(M)$ . Тогда для рассмотренных выше гиперэкспоненциальных распределений получим:

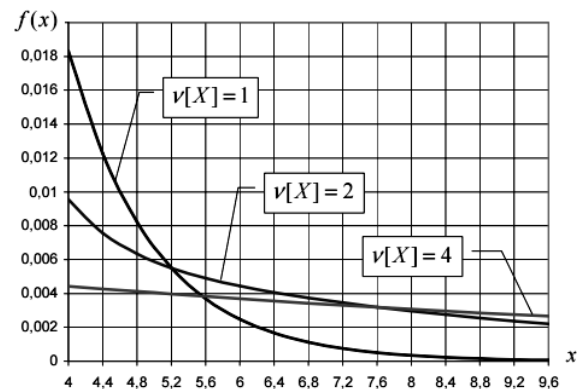
$$Pr(X < M) = F(M) = \begin{cases} 0,735 & \text{для распределения с } \nu[X] = 2; \\ 0,919 & \text{для распределения с } \nu[X] = 4. \end{cases}$$

Таким образом, более 73% значений случайной величины, распределенной по гиперэксп закону с коэффициентом вариации, равным 2, попадает в интервал  $(0; M)$  и только 27% значений окажутся больше математического ожидания. Для случайной величины, с коэффициентом вариации, равным 4, вероятность попадания в интервал  $(0; M)$  еще выше и составляет почти 92%. Очевидно, что чем больше коэффициент вариации, тем больше вероятность появления маленьких значений случайной величины.



Про «Хвост» распределения: На графике для СВ  $> 4$ .

Из графика видно, что кривая плотности гиперэксп распределения с коэффициентом вариации, равным 4, имеет длинный так называемый «тяжелый хвост», характеризующийся малым изменением. Это означает, что при гиперэксп распределении вероятность появления больших значений случайной величины значительно выше, чем, например, для эксп распределения. Таким образом, основное отличие гиперэксп распределения от эксп состоит в том, что гиперэксп распределение характеризуется большей вероятностью появления маленьких значений СВ и, в то же время, большей вероятностью появления больших значений СВ.



## 6. Гиперэрланговское распределение

Гиперэрл распределение представляет собой аддитивную смесь нормированных распределений Эрланга и является наиболее общим распределением неотрицательных непрерывных СВ, поскольку имеет **коэффициент вариации в интервале от 0 до  $\infty$** . Составляющими гиперэрл распределения, в отличие от гиперэксп, являются нормированные распределения Эрланга.

$$f(x) = \sum_{i=1}^n q_i \frac{k_i \alpha_i (k_i \alpha_i x)^{k_i-1}}{(k_i-1)!} e^{-k_i \alpha_i x} \quad (x \geq 0)$$

Числовые характеристики распределений

Распределение	$M[X]$	$\alpha_2[X]$	$D[X]$	$\sigma[X]$	$\nu[X]$	Примечания
Пуассона	$a$	$a(a+1)$	$a$	$\sqrt{a}$	$1/\sqrt{a}$	$a > 0$
Геометрическое	$\frac{1-\gamma}{\gamma}$	$\frac{2(1-\gamma)^2}{\gamma^2}$	$\frac{(1-\gamma)^2}{\gamma^2}$	$\frac{1-\gamma}{\gamma}$	1	$0 < \gamma < 1$
Равномерное	$\frac{a+b}{2}$	$\frac{a^2+ab+b^2}{3}$	$\frac{(b-a)^2}{12}$	$\frac{b-a}{2\sqrt{3}}$	$\frac{b-a}{\sqrt{3}(a+b)}$	$b > a$
Экспоненциальное	$\frac{1}{\alpha}$	$\frac{2}{\alpha^2}$	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha}$	1	$\alpha > 0$
Эрланга	$\frac{k}{\alpha}$	$\frac{k(k+1)}{\alpha^2}$	$\frac{k}{\alpha^2}$	$\frac{\sqrt{k}}{\alpha}$	$\frac{1}{\sqrt{k}}$	$k = 1, 2, \dots$
Эрланга нормированное	$\frac{1}{\alpha}$	$\frac{k+1}{k\alpha^2}$	$\frac{1}{k\alpha^2}$	$\frac{1}{\alpha\sqrt{k}}$	$\frac{1}{\sqrt{k}}$	$k = 1, 2, \dots$
Гиперэкспоненциальное	$\sum_{i=1}^n \frac{q_i}{\alpha_i}$	$2 \sum_{i=1}^n \frac{q_i}{\alpha_i^2}$	$\alpha_2[X] - (M[X])^2$	$\sqrt{D[X]}$	$\nu[X] \geq 1$	$\sum_{i=1}^n q_i = 1$ $\alpha_i > 0$
Гиперэрланговское	$\sum_{i=1}^n \frac{q_i}{\alpha_i}$	$\sum_{i=1}^n q_i \frac{k_i+1}{k_i \alpha_i^2}$	$\alpha_2[X] - (M[X])^2$	$\sqrt{D[X]}$	$\nu[X] \geq 0$	$\sum_{i=1}^n q_i = 1$ $\alpha_i > 0$ $k_i = 1, 2, \dots$

- математическое ожидание  $M[X]$ ;
- второй начальный момент  $\alpha_2[X]$ ;
- дисперсия  $D[X]$ ;
- среднее квадратическое отклонение  $\sigma[X]$ ;
- коэффициент вариации  $\nu[X]$ .

$$D[X] = \alpha_2[X] - (M[X])^2;$$

$$\sigma[X] = \sqrt{D[X]};$$

$$\nu[X] = \frac{\sigma[X]}{M[X]}.$$

## 12. Приближенное сведение немарковских процессов к Марковским. Метод псевдосостояний.

Реальные процессы очень часто обладают тем или другим последствием. Для Марковского процесса время пребывания системы подряд в каком-либо состоянии распределено по показательному закону, но это не всегда бывает так. Например, если поток событий, приводящий систему из состояния в состояние есть поток отказов какого-то узла, то более естественно предположить, что оставшееся время безотказной работы узла *зависит от того, сколько времени узел работал*. При этом время пребывания узла в рабочем состоянии представляет собой СВ, распределённую не по показательному закону. Возникает вопрос, можно ли приближенно заменять непуассон потоки – пуассоновскими и к каким ошибкам в предельных вероятностях состояний может привести подобная замена.

Пусть в некоторую физическую систему  $S$ , в которой протекает случайный процесс, направляем какими-то непуассоновскими потоками событий. Если попробовать записать ДУ, выражающие вероятности состояний как функции времени, увидим, что в общем случае это не удастся. Например, для Марковского процесса мы вычисляли вероятность того, что в момент  $t + \Delta t$  система будет в состоянии  $S_i$ , учитывая только то, *в каком состоянии* система была в момент  $t$ , и не учитывая, *сколько времени* она была в этом состоянии. Для немарковской системы этот прием уже непригоден: вычисляя вероятность перехода из состояния в другое за время  $\Delta t$ , мы должны учитывать, сколько времени система *уже провела* в данном состоянии. Это приводит, вместо обыкновенных дифференциальных уравнений, к уравнениям с частными производными, то есть к гораздо более сложному математическому аппарату.

Возникает вопрос: можно ли искусственно свести немарковский процесс к Марковскому. Если число состояний системы не очень велико, а отличающиеся от простейших потоки событий представляют собой потоки Эрланга. Тогда, вводя в схему возможных состояний системы некоторые фиктивные «псевдосостояния», удастся свести немарковский процесс к Марковскому и описать его с помощью обыкновенных ДУ, которые при  $t \rightarrow \infty$  переводят в алгебраические уравнения для предельных вероятностей состояний.

**Пример.** Пусть некоторая система, которая под влиянием некоторого простейшего потока может выходить из строя с интенсивностью  $\lambda$ .

Отказавшее устройство немедленно начинает восстанавливаться и восстанавливается за время  $T$ , распределенного по непоказательному закону, а по закону Эрланга 3-его порядка:

$$f_3(t) = \frac{\mu (\mu t)^2}{2} e^{-\mu t} \quad (t \geq 0)$$

,  $\mu$  – интенсивность восстановления.

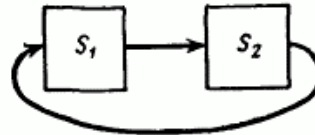


Требуется свести данный немарковский процесс к Марковскому и найти для него предельные состояния вероятностей.

**Решение:** т.к интервал времени  $T$  между соседними событиями в потоке Эрланга  $k$ -ого порядка представляет собой сумму  $k$  независимых СВ – расстояний между событиями в исходном простейшем потоке  $\rightarrow T = T_1 + T_2 + T_3$ , каждая из которых распределена по показательному закону с параметром  $\mu$ .

$$f_1(t) = \mu e^{-\mu t} \quad (t > 0).$$

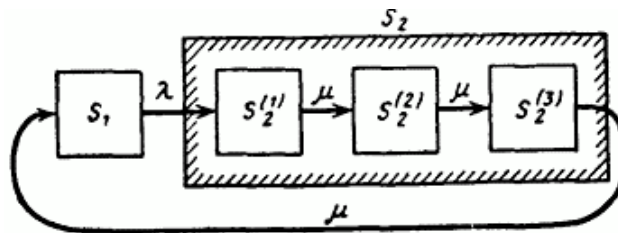
А истинных состояний у системы всего 2:  $S_1$  – устройство исправно,  $S_2$  – устройство ремонтируется.



Однако в виду того, что переход по стрелке  $S_2 \rightarrow S_1$  происходит под влиянием не простейшего, а эрланговского потока событий, процесс, происходящий в системе, Марковским не является, и для него мы не можем написать ни ДУ, не АУ. Чтобы искусственно свести этот процесс к Марковскому, введем в цепочку состояний, вместо состояния  $S_2$ , три последовательных «псевдосостояния»:

$S_2^{(1)}$  – ремонт начинается;  
 $S_2^{(2)}$  – ремонт продолжается;  
 $S_2^{(3)}$  – ремонт заканчивается,

Т.е разделим ремонт на три этапа, причем время пребывания системы в каждом по показательному закону. Процесс протекающий в систему будет уже Марковски



Вероятности пребывания в псевдосостояниях соответственно равны  $p_2^{(1)}$ ,  $p_2^{(2)}$ ,  $p_2^{(3)}$ .

$$p_2 = p_2^{(1)} + p_2^{(2)} + p_2^{(3)}. \quad \text{Обозначая } \bar{t}_1 = 1/\lambda, \quad \bar{t}_2 = 1/\mu,$$

Можем сразу написать предельные вероятности состояний:

$$\left. \begin{aligned} p_1 &= \frac{\bar{t}_1}{\bar{t}_1 + \bar{t}_2 + \bar{t}_2 + \bar{t}_2} = \frac{\bar{t}_1}{\bar{t}_1 + 3\bar{t}_2}; \\ p_2^{(1)} &= p_2^{(2)} = p_2^{(3)} = \frac{\bar{t}_2}{\bar{t}_1 + 3\bar{t}_2}; \\ p_2 &= p_2^{(1)} + p_2^{(2)} + p_2^{(3)} = \frac{3\bar{t}_2}{\bar{t}_1 + 3\bar{t}_2}. \end{aligned} \right\}$$

Заметим, что величина  $3\bar{t}_2$  представляет собой не что иное, как среднее время восстановления – оно равно сумме средним времен пребывания системы в

каждой фазе ремонта.

Переходя в формулах для  $p$  от средних времен  $t$  к интенсивностям потоков, по формулам ... получим...:

$$\bar{t}_1 = 1/\lambda, \quad \bar{t}_2 = 1/\mu \qquad \rho_1 = \mu/(\mu + 3\lambda), \quad \rho_2 = 3\lambda/(\mu + 3\lambda).$$

Таким образом, вероятность пребывания в каждом из двух состояний, как и для Марковского цикла, равна относительному среднему времени пребывания подряд в каждом из состояний.

\*Если что не понятно: [http://stu.sernam.ru/book\\_rop.php?id=49](http://stu.sernam.ru/book_rop.php?id=49)

## **\*\* . Обозначения в СМО. Формулы Литтла.**

Основные элементы СМО:

$a$  – вид распределения моментов поступления заявок на обслуживания (входной поток)

$b$  – вид распределения времени обслуживания (выходной поток)

$c$  – число параллельно функционирующих каналов

$d$  – порядок обслуживания (ПЕРППО, ПОСППО, СОЗ)

$e$  – максимальное число допускаемых в системе заявок ( $n+m$ )

$f$  – емкость источника, генерирующего заявки на обслуживание.

Для  $a$  и  $b$  применяются следующие обозначения:

M-пуассоновское распределение заявок (простейший поток)

D – детерминированный интервал времени между событиями потока

E $k$ - распределение эрланга  $k$ -го порядка

GI-любое (произвольное) распределение времени между событиями входного потока.

G любое распределение времени между событиями выходного потока

GD – (general discipline) – любая дисциплина очереди (т.е. формулы не зависят от вида дисциплины очереди).

**Обозначения Кендалла (1953г):**

**(a/b/c)**

**Обозначения А.М.Ли. В 1966г.он добавил символы  $d$  и  $e$ , а X.Taxa добавил  $f$**

**(a/b/c):(d/e/f).**

**Пример. (M/D/10):(GD/N/∞).**

Это модель СМО с пуассоновским входным потоком, фиксированным постоянным временем обслуживания и 10 параллельно работающими каналами обслуживания. Дисциплина очереди не регламентирована. Данная система не может вместить более N заявок (очередь + обслуживаемые клиенты). Источник, порождающий заявки, имеет бесконечную ёмкость.

Буквенные обозначения в американской литературе:

$P_n$  – вероятность того, что в системе находится  $n$  клиентов (заявок),

$L_s$  – среднее число находящихся в системе клиентов (заявок на обслуживание),

$L_q$  – среднее число клиентов в очереди на обслуживание,



$W_s$  – средняя продолжительность пребывания клиента в *системе*,

$W_q$  - средняя продолжительность пребывания клиента в *очереди*.

По определению

$$L_s = \sum_{n=1}^{\infty} np_n$$

$$L_q = \sum_{n=c}^{\infty} (n-c)p_n$$

Между параметрами  $L$  и  $W$  существует фундаментальная связь, определяемая формулами Литтла

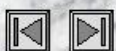
$$L_s = \lambda W_s ;$$

$$L_q = \lambda W_q .$$

При этом не накладывается никаких ограничений на распределения входного потока и времени обслуживания. Впервые доказательство этого факта дал Дж.Литтл и это соотношение носит название **формула Литтла**. Эти формулы справедливы для систем без отказов. В системах с отказами нужно эти величины умножить на  $(1 - p_{отк}) = q$  – относительную пропускную способность.

## Формула Литтла 1

для любой СМО, при любом характере потока заявок, при любом распределении времени обслуживания, при любой дисциплине обслуживания **среднее время пребывания заявки в системе равно среднему числу заявок в системе, деленному на интенсивность потока заявок**



## Сети массового обслуживания

**Сеть массового обслуживания (СеМО)** – совокупность взаимосвязанных СМО, в среде которых циркулируют заявки (рис.3.2,а).

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

**Узел** сети представляет собой систему массового обслуживания.

**Источник** – генератор заявок, поступающих в сеть и требующих определенных этапов обслуживания в узлах сети.

Для упрощенного изображения СеМО используется граф СеМО.

**Граф СеМО** – ориентированный граф, вершины которого соответствуют узлам СеМО, а дуги отображают переходы заявок между узлами (рис.3.2,б).

Переходы заявок между узлами СеМО, в общем случае, могут быть заданы в виде вероятностей передач.

Путь движения заявок в СеМО называется *маршрутом*.

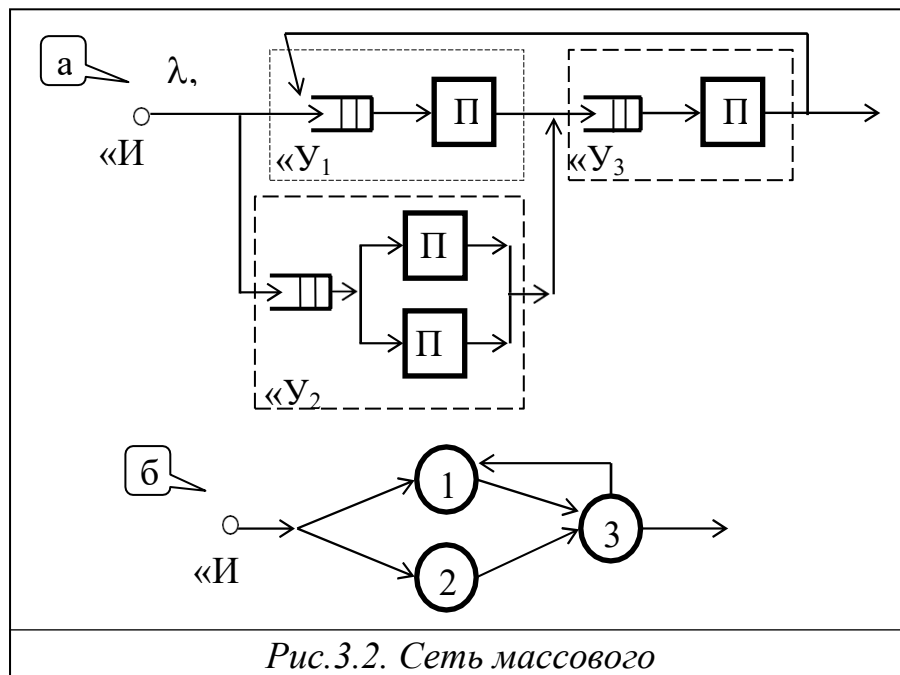
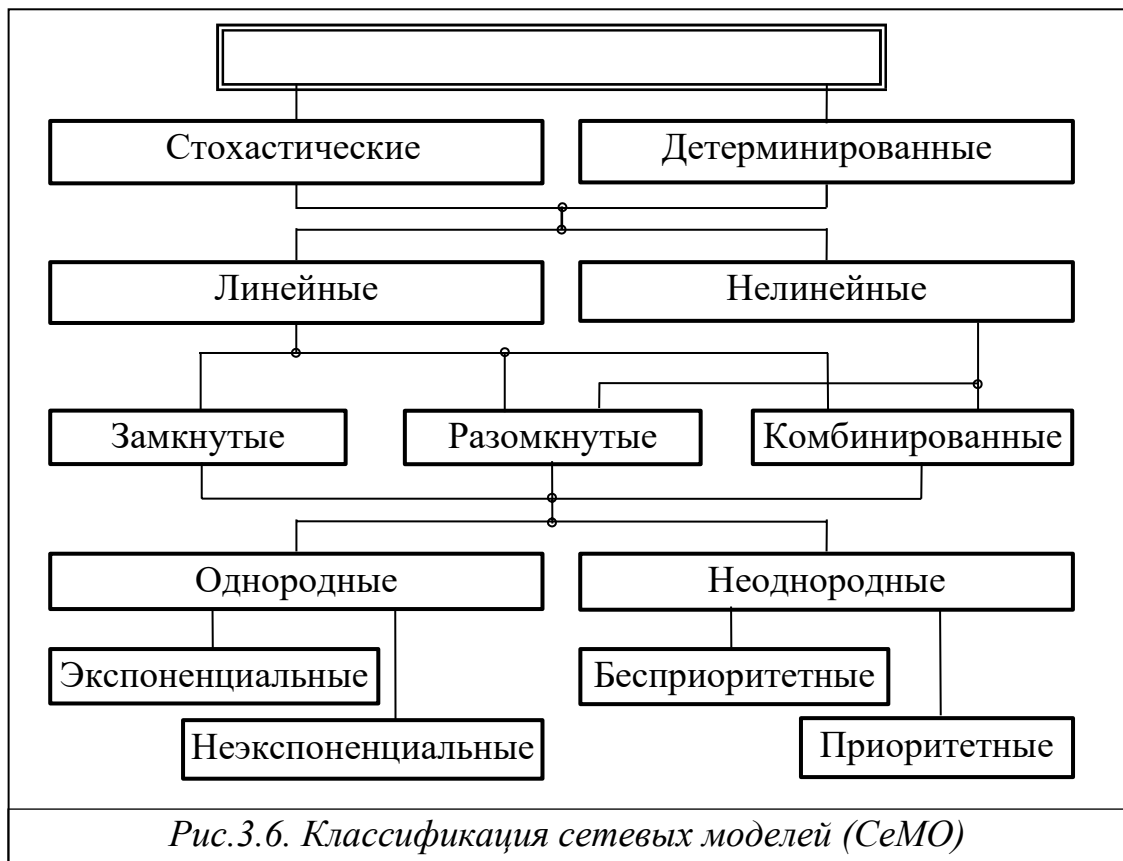


Рис.3.2. Сеть массового

### Сетевые модели

В зависимости от структуры и свойств исследуемых систем их моделями могут служить СеМО различных классов. Одна из возможных классификаций сетевых моделей приведена на рис.3.6.



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

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

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

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

В *линейных* СеМО, как это следует из определения, интенсивность потока заявок в узел  $j$  связана с интенсивностью потока заявок в узел  $i$  линейной зависимостью:

$$\lambda_j = \alpha_{ij} \lambda_i,$$

где  $\alpha_{ij}$  – коэффициент пропорциональности, показывающий, во сколько раз отличаются интенсивности потоков заявок в узел  $j$  и в узел  $i$  ( $i, j = \overline{1, n}$ ).

Поскольку указанная зависимость справедлива для любой пары узлов, это выражение можно записать в несколько ином виде и выразить интенсивность поступления заявок во все узлы  $j = \overline{1, n}$  через одну и ту же интенсивность, например, через интенсивность  $\lambda_0$  потока заявок, поступающих в СеМО из

источника заявок:

$$\lambda_j = \alpha_j \lambda_0.$$

В последнем выражении коэффициент пропорциональности  $\alpha_j \geq 0$  показывает, во сколько раз интенсивность потока заявок в узел  $j$  ( $i, j = \overline{1, n}$ ) отличается от интенсивности источника заявок, и называется коэффициентом передачи. Коэффициент передачи может принимать любое положительное значение.

Коэффициент передачи играет важную роль при разработке математических зависимостей и расчете характеристик функционирования сетевых моделей. Это обусловлено тем физическим смыслом, который несет в себе коэффициент передачи.

Коэффициент передачи можно трактовать как среднее число попаданий заявки в данный узел за время ее нахождения в сети. Например, если коэффициент передачи узла СеМО равен 3, то это означает, что любая заявка за время нахождения в сети в среднем 3 раза побывает на обслуживании в данном узле. Значение коэффициента передачи, равное 0,25, будет означать, что в среднем только одна заявка из четырёх попадёт на обслуживание в данный узел, а три другие обойдут данный узел стороной.

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

*Нелинейность СеМО* может быть обусловлена:

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

Таким образом, СеМО является линейной, если в ней заявки не размножаются и не теряются. Ниже рассматриваются, в основном, линейные СеМО.

3. По числу циркулирующих в сети заявок различают СеМО:

- разомкнутые;
- замкнутые;
- замкнуто-разомкнутые.

**Разомкнутая (открытая) СеМО (РСеМО)** содержит один или несколько *внешних независимых источников* заявок, которые генерируют заявки в сеть независимо от числа заявок, находящихся в сети (рис.3.7,а). В РСеМО одновременно может находиться *любое число заявок*, в том числе, и сколь угодно большое, то есть от 0 до бесконечности. С РСеМО связана внешняя среда, из которой поступают заявки в сеть и в которую они возвращаются после обслуживания в сети. Внешняя среда в РСеМО обозначается обычно как нулевой узел "0", и РСеМО, в этом случае, изображается в виде рис.3.7,б.

**Замкнутая (закрытая) СеМО (ЗСеМО)** не содержит *независимых внешних источников* заявок и характеризуется тем, что в ней циркулирует *постоянное число заявок*  $M$  (рис.3.7,в). На графе ЗСеМО из физических соображений,

связанных с конкретным представлением процесса функционирования исследуемой реальной системы, обычно выделяется особая дуга, отображающая процесс завершения обслуживания заявок в сети и мгновенного формирования новой заявки с такими же параметрами обслуживания, что и завершившая обслуживание. Такая трактовка позволяет рассматривать завершившую обслуживание заявку как новую заявку, поступившую в сеть из *зависимого источника* заявок.

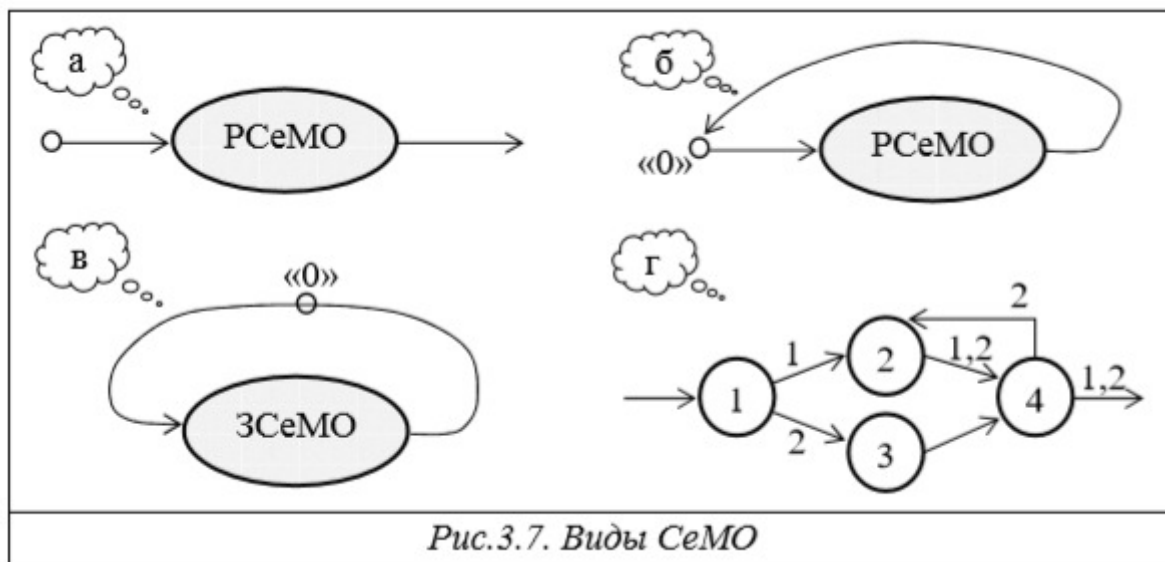


Рис.3.7. Виды СеМО

По аналогии с PCeMO на выделенной дуге ЗСеМО отмечается условная точка "0", рассматриваемая как нулевой узел и трактуемая иногда как фиктивная СМО с нулевой длительностью обслуживания или как зависимый источник заявок, генерирующий заявки только в момент поступления некоторой заявки на его вход. Выделение нулевого узла в ЗСеМО преследует двоякую цель: во-первых, достигается однозначность в представлении и математическом описании PCeMO и ЗСеМО; во-вторых, обеспечивается возможность определения временных характеристик ЗСеМО относительно выделенного узла "0". В частности, **время пребывания заявок в ЗСеМО** рассматривается как промежуток времени между двумя соседними моментами прохождения заявки через нулевой узел.

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

4. По типу циркулирующих заявок различают СеМО:

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

Маршруты заявок разных классов задаются путем указания номеров классов заявок на соответствующих дугах сети (рис.3.7,г).

### 3.1. Параметры и характеристики СеМО

#### • Параметры СеМО

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

- число узлов в сети:  $n$ ;
- число обслуживающих приборов в узлах сети:  $K_1, \dots, K_n$ ;
- матрица вероятностей передач:  $P = [p_{ij} \mid i, j = 0, 1, \dots, n]$ , где  $p_{ij}$  – вероятность передачи заявки из узла  $i$  в узел  $j$ ;
- интенсивность  $\lambda_0$  источника заявок, поступающих в **разомкнутую** СеМО (РСеМО), или число заявок  $M$ , циркулирующих в **замкнутой** СеМО (ЗСеМО);
- средние длительности обслуживания заявок в узлах сети:  $b_1, \dots, b_n$

Заметим, что состав параметров разомкнутых и замкнутых СеМО различается только одним параметром, а именно: для ЗСеМО, в отличие от РСеМО, вместо интенсивности  $\lambda_0$  поступления заявок в сеть необходимо задать число постоянно циркулирующих в сети заявок  $M$ .

Для *линейных* СеМО элементы матрицы вероятностей передач должны удовлетворять условию:

$$\sum_{j=0} p_{ij} = 1 \quad (i = \overline{0, n}) \quad (3.24)$$

Это условие отражает тот факт, что любая заявка, покинувшая некоторый узел, обязательно (с вероятностью 1) перейдёт в какой-то узел, включая тот же самый или нулевой. Переход заявки в нулевой узел означает, что заявка покинула сеть.

В случае *неэкспоненциальных* разомкнутых СеМО дополнительно необходимо задать законы распределения или, по крайней мере, вторые моменты интервалов времени между поступающими в разомкнутую сеть заявками и длительностей обслуживания заявок в узлах сети.

В случае *неоднородных* СеМО необходимо дополнительно задать количество классов заявок  $H$  в сети и для каждого класса – матрицы вероятностей передач  $P(h)$ , интенсивности  $\lambda_0(h)$  или число заявок  $M(h)$ , а также средние длительности обслуживания  $b_i(h)$  заявок класса  $h = \overline{1, H}$  в узле  $i = \overline{1, n}$ . При необходимости могут быть заданы законы распределений интервалов между поступающими в РСеМО заявками и законы распределений длительностей обслуживания заявок разных классов в узлах сети.

#### Режимы функционирования СеМО

СеМО, как и СМО, может работать в установившемся и неуставившемся режимах. Последний может быть связан с началом работы системы (переходной режим), нестационарным характером потока заявок и обслуживания в приборе (нестационарный режим) и перегрузкой системы (режим перегрузки).

Очевидно, что для СеМО, как и для СМО, при использовании предположения о стационарности входящего потока заявок и длительностей обслуживания заявок в узлах *условие существования установившегося режима совпадает с условием отсутствия перегрузок*.

Рассмотрим это условие для разомкнутой и замкнутой СеМО.

Очевидно, что перегрузки в **разомкнутой СеМО** отсутствуют, если каждый узел сети работает без перегрузок. Если же хотя бы один из узлов сети не справляется с нагрузкой, то длина очереди в этом узле начнет увеличиваться до бесконечности и, следовательно, суммарное число заявок в РСеМО будет расти неограниченно.

Таким образом, для того чтобы в *разомкнутой СеМО* не было перегрузок, необходимо отсутствие перегрузок во всех узлах РСеМО, то есть загрузка  $\rho_j$  любого узла  $j$  ( $j = \overline{1, n}$ ) должна быть строго меньше единицы:

$$\rho_j = \frac{\lambda_j b_j}{K_j} = \frac{\alpha_j \lambda_0 b_j}{K_j} < 1 \quad \text{для всех } (j = \overline{1, n}).$$

Из последнего неравенства имеем:

$$\lambda_0 < \frac{K_j}{\alpha_j b_j} \quad \text{для всех } (j = \overline{1, n}).$$

Это условие может быть записано также в следующем виде:

$$\lambda_0 < \min \left( \frac{K_1}{\alpha_1 b_1}, \frac{K_2}{\alpha_2 b_2}, \frac{K_3}{\alpha_3 b_3}, \frac{K_4}{\alpha_4 b_4} \right). \quad (3.25)$$

Полученное условие налагает ограничение сверху на интенсивность поступления заявок в РСеМО из внешнего источника. Узлы, в которых указанное условие не выполняется, являются перегруженными. С течением времени это приводит к неограниченному росту числа заявок в сети, которые скапливаются в перегруженных узлах, имеющих накопители неограниченной ёмкости.

В дальнейшем при исследовании разомкнутых СеМО, если не оговорено другое, будем полагать, что в сети существует установившийся режим.

Несколько иначе дело обстоит для **замкнутых СеМО**. Поскольку в ЗСеМО циркулирует постоянное число заявок, то в узлах сети не могут образовываться очереди бесконечной длины, следовательно, в ЗСеМО всегда существует установившийся режим. Даже если в сети имеется очень «медленный» узел, в котором по сравнению с другими узлами слишком долго обрабатываются заявки, то это может привести только к тому, что все заявки будут постоянно скапливаться в очереди перед данным узлом, однако их количество будет всегда конечно и в пределе равно числу циркулирующих в сети заявок. Загрузка такого «медленного» узла будет близка к единице, поскольку постоянное наличие очереди перед этим узлом обуславливает непрерывную работу приборов узла. Такой узел обычно представляет собой так называемое «узкое место» сети.

#### ● Характеристики СеМО

Характеристики СеМО делятся на два класса:

- **узловые**, описывающие эффективность функционирования отдельных

узлов СеМО;

- **сетевые**, описывающие функционирование СеМО в целом. Состав *узловых характеристик* СеМО, работающей в *стационарном режиме*, такой же, как и для СМО, и для узла  $j = \overline{1, n}$  включается в себя следующие характеристики:

- *нагрузка узла*:  $y_j = \lambda_j b_j = \alpha_j \lambda_0 b_j$ ;
- *загрузка узла*:  $\rho_j = \frac{y_j}{K_j} = \frac{\alpha_j \lambda_0 b_j}{K_j}$ , причем  $\rho_j < 1$ ;
- *коэффициент простоя узла*:  $\eta_j = 1 - \rho_j$ ;
- *время ожидания заявок в узле*:  $w_j$ ;
- *время пребывания заявок в узле*:  $u_j = w_j + b_j$ ;
- *длина очереди заявок в узле*:  $l_j = \lambda_j w_j = \alpha_j \lambda_0 w_j$ ;
- *число заявок в узле (в очереди и на обслуживании)*:  
 $m_j = \lambda_j u_j = \alpha_j \lambda_0 (w_j + b_j) = l_j + y_j$ .

В приведенных выше формулах использован тот факт, что в линейных СеМО интенсивность поступления заявок в любой узел связана с интенсивностью источника соотношением (3.5).

На основе *узловых характеристик* рассчитываются *сетевые характеристики* СеМО:

- **суммарная нагрузка** во всех узлах, характеризующая *среднее число заявок, одновременно находящихся на обслуживании во всех узлах сети*:

$$Y = \sum_{j=1}^n y_j ,$$

где  $y_j$  – нагрузка узла  $j$ , причём  $0 < Y < \sum_{j=1}^n K_j$ ;

- **суммарная загрузка** всех узлов СеМО, характеризующая *среднее число параллельно работающих узлов сети*:

$$R = \sum_{j=1}^n \rho_j ,$$

где загрузка узла  $j$ , причём  $0 < R \leq n$ ;

- *среднее число заявок, находящихся в очередях всех узлов сети и ожидающих обслуживания*:

$$L = \sum_{j=1}^n l_j , \tag{3.26}$$

где  $l_j$  – средняя длина очереди заявок в узле  $j$ ;

- *среднее число заявок, находящихся в сети*:



$$M = \sum_{j=1}^n m_j, \quad (3.27)$$

где  $m_j$  – среднее число заявок в узле  $j$ , причём для замкнутых сетей это выражение может быть использовано для проверки правильности проведенных расчетов, так как для них число заявок  $M$  в сети задано;

- среднее время ожидания заявок в сети:

$$W = \sum_{j=1}^n \alpha_j w_j, \quad (3.28)$$

где  $w_j$  – среднее время ожидания заявок в узле  $j$ ;  $\alpha_j$  – коэффициент передачи для узла  $j$ , оказывающий среднее число попаданий заявки в узел  $j$  за время её нахождения в сети;  $W_j = \alpha_j w_j$  – представляет собой суммарное (полное) время ожидания заявки в узле  $j$  за время её нахождения в сети;

- среднее время пребывания заявок в сети:

$$U = \sum_{j=1}^n \alpha_j u_j, \quad (3.29)$$

где  $u_j$  – среднее время пребывания заявок в узле  $j$ .  $U_j = \alpha_j u_j$  – суммарное (полное) время пребывания заявки в узле  $j$  за время её нахождения в сети;

- **производительность замкнутой СеМО**  $\lambda_0$ , определяемая как интенсивность потока заявок, проходящих через выделенный нулевой узел замкнутой сети, и представляющая собой среднее число заявок, обслуженных в ЗСеМО за единицу времени; производительность ЗСеМО может быть рассчитана на основе выражения (3.5), из которого следует:

$$\lambda_0 = \lambda_j / \alpha_j \quad (j = 1, \dots, n); \quad (3.30)$$

Следует отметить, что для сетевых характеристик СеМО выполняются те же фундаментальные соотношения, что и для СМО, а именно:

$$L = \lambda_0 W; \quad (3.31)$$

$$M = \lambda_0 U; \quad (3.32)$$

$$M = L + Y; \quad (3.33)$$

$$U = W + B, \quad (3.34)$$

где  $B = \sum_{j=1}^n \alpha_j b_j$  – суммарное время обслуживания заявки во всех узлах за время её нахождения в сети.

Выражения (3.31) и (3.32) представляют собой формулы Литтла для расчёта сетевых характеристик СеМО.

Из (3.32) может быть получена ещё одна важная формула для расчёта производительности ЗСеМО:

$$\lambda_0 = \frac{M}{U}. \quad (3.35)$$

Для неоднородной СеМО перечисленные характеристики определяются как для каждого класса в отдельности, так и для объединенного (суммарного) потока заявок.

### 3.1. Разомкнутые экспоненциальные СеМО с однородным потоком заявок

#### • Описание разомкнутых СеМО

Рассмотрим разомкнутую экспоненциальную сеть массового обслуживания (СеМО) с однородным потоком заявок при следующих предположениях:

- разомкнутая СеМО (РСеМО) *произвольной топологии* содержит  $n$  узлов;
- после завершения обслуживания в каком-либо узле передача заявки в другой узел происходит *мгновенно*;
- в качестве узлов могут быть как *одноканальные*, так и *многоканальные* СМО;
- все приборы многоканального узла являются *идентичными*, и любая заявка может обслуживаться любым прибором;
- заявка, поступившая в многоканальный узел, когда все или несколько приборов свободны, направляется случайным образом в *любой свободный прибор*;
- в каждом узле РСеМО имеется накопитель заявок *неограниченной ёмкости*, что означает отсутствие отказов поступающим заявкам при их постановке в очередь, то есть любая поступающая в узел заявка всегда найдет в накопителе место для ожидания независимо от того, сколько заявок уже находится в очереди;
- заявки поступают в РСеМО из *внешнего независимого* источника и образуют *простейший поток* заявок;
- длительности обслуживания заявок во всех узлах сети представляют собой случайные величины, распределенные по *экспоненциальному закону*;
- обслуживающий прибор любого узла *не простаивает*, если в его накопителе имеется хотя бы одна заявка, причем после завершения обслуживания очередной заявки *мгновенно* из накопителя выбирается следующая заявка;
- в каждом узле сети заявки из накопителя выбираются в соответствии с *бесприоритетной* дисциплиной обслуживания в порядке поступления (ОПП) по правилу «первым пришел – первым обслужен» (FIFO – First In First Out).

Для описания *линейных разомкнутых однородных экспоненциальных* СеМО необходимо задать следующую совокупность **параметров**:

- *число узлов* в сети:  $n$ ;
- *число обслуживающих приборов* в узлах сети:  $K_1, \dots, K_n$ ;
- *матрицу вероятностей передач*:  $\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$ , где вероятности передач  $p_{ij}$  должны удовлетворять условию (3.23); сумма элементов каждой строки должна быть равна 1;
- *интенсивность*  $\lambda_0$  источника заявок, поступающих в РСеМО;

- средние *длительности* обслуживания заявок в узлах сети:  $b_1, \dots, b_n$ .

На основе перечисленных параметров могут быть рассчитаны узловые и сетевые характеристики, описывающие эффективность функционирования соответственно узлов и РСемо в целом.

Расчет характеристик функционирования *линейных разомкнутых однородных экспоненциальных* СеМО базируется на эквивалентном преобразовании сети и проводится в четыре этапа:

- расчет коэффициентов передач  $\alpha_j$  и интенсивностей потоков заявок  $\lambda_j$  в узлах ( $j = 1, \dots, n$ ) СеМО;
- проверка условия отсутствия перегрузок в СеМО;
- расчет узловых характеристик;
- расчет сетевых характеристик.

### 1) Расчет коэффициентов передач и интенсивностей потоков заявок в узлах РСемо

Покажем, что интенсивности  $\lambda_0, K, \lambda_n$  потоков заявок, поступающих в узлы  $0, \dots, n$  сети, однозначно определяются вероятностями передач  $p_{ij}$  ( $i, j = 1, \dots, n$ ), задающими маршруты заявок в СеМО.

Будем рассматривать только установившийся режим.

Так как в линейной СеМО заявки не размножаются и не теряются, то интенсивности входящего и исходящего потоков для любого узла будут равны между собой.

Интенсивность потока заявок, входящих в любой узел  $j$  сети, равна сумме интенсивностей потоков заявок, поступающих в него из других узлов  $i = \overline{0, n}$  (рис 4.11). Поскольку заявки из узла  $i$  поступают в узел  $j$  с вероятностью  $p_{ij}$ , то интенсивность потока заявок, поступающих из  $i$  в  $j$ , равна  $p_{ij} \lambda_i$ ,  $\lambda_i$  интенсивность исходящего и, следовательно, входящего потока заявок узла  $i$ . С учетом этого, на входе узла  $j$  имеется поток с интенсивностью:

$$\lambda_j = \sum_{i=0}^n p_{ij} \lambda_i \quad (i = 0, 1, \dots, n). \quad (4.16)$$

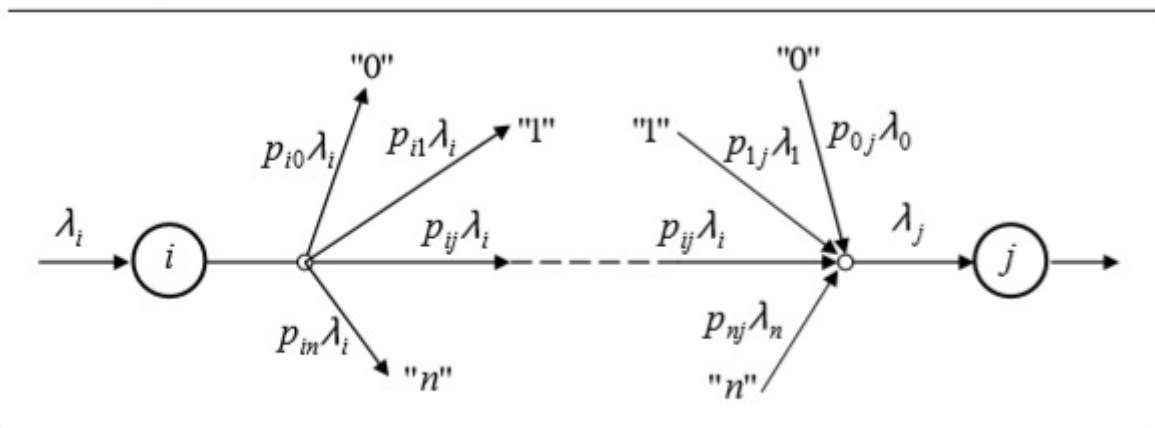


Рис.4.11. К расчёту интенсивностей потоков заявок в узлах РСемо

Выражение (4.16) представляет собой *систему линейных алгебраических уравнений*  $(n + 1)$ -го порядка, из которой могут быть найдены интенсивности потоков заявок в виде соотношения  $\lambda_j = \alpha_j \lambda_0$  ( $j = 1, \dots, n$ ). Коэффициент  $\alpha_j$

называется **коэффициентом передачи** и определяет среднее число попаданий заявки в узел  $j$  за время ее нахождения в сети, причем  $\alpha_0 = 1$ .

Для разомкнутой СеМО известна интенсивность источника заявок  $\lambda_0$ . Можно показать, что система уравнений для расчета интенсивностей имеет единственное решение вида  $\lambda_j = \alpha_j \lambda_0$ , где  $\lambda_0$  – заданная величина.

- Проверка условия отсутствия перегрузок в СеМО

В п.3.4.2 показано, что в разомкнутой СеМО отсутствуют перегрузки, если выполняется условие (3.25):

$$\lambda_0 < \min \left( \frac{K_1}{\alpha_1 b_1}, \frac{K_2}{\alpha_2 b_2}, \frac{K_3}{\alpha_3 b_3}, \frac{K_4}{\alpha_4 b_4} \right).$$

Если указанное условие не выполняется, то, как следует из него, стационарный режим в разомкнутой СеМО может быть реализован одним из следующих способов:

- уменьшением интенсивности  $\lambda_0$  внешнего источника заявок до значения, при котором это условие будет выполняться;
- увеличением количества обслуживающих приборов  $K_j$  в перегруженных узлах;
- уменьшением длительностей  $b_j$  обслуживания заявок в перегруженных узлах.
- Уменьшением коэффициентов передач  $\alpha_j$  в перегруженных узлах.
- 

- Расчет узловых характеристик РСемо

Один и тот же объект, рассматриваемый на разных уровнях детализации, можно представить различными моделями массового обслуживания, характеристики которых одинаковы или отличаются на величину, не превосходящую заданной погрешности. При выполнении определенных условий такие модели легко преобразуются друг в друга.

Для сетевых моделей в виде разомкнутых и замкнутых СеМО могут использоваться два вида преобразований:

- эквивалентное преобразование;
- толерантное преобразование.

Две сетевые модели *эквивалентны*, если сравниваемые характеристики этих моделей не отличаются друг от друга.

Две сетевые модели *толерантны* (подобны), если значения определенных характеристик отличаются друг от друга на величину, не превосходящую заданную.

Использование свойств эквивалентных и толерантных моделей позволяет упростить расчет характеристик моделей путем замены сложных сетевых моделей более простыми. Эквивалентными могут быть сетевые модели одного типа (например, две замкнутые сети), толерантными — модели как одного, так и разных типов [11].

Расчет характеристик функционирования линейных разомкнутых однородных экспоненциальных СеМО базируется на эквивалентном преобразовании сети, заключающемся в представлении *разомкнутой СеМО с  $n$*

узлами в виде  $n$  независимых экспоненциальных СМО типа М/М/Ν (простейший поток заявок, длительность обслуживания распределена по экспоненциальному закону,  $N$  обслуживающих приборов). При этом интенсивность входящего потока заявок в СМО, отображающую узел ( $j=1,...,n$ ) сети, определяется из системы алгебраических уравнений (4.16) через интенсивность входящего в сеть потока и коэффициент передачи узла:  $\lambda_j = \alpha_j \lambda_0$ , а средняя длительность обслуживания заявок в СМО равна длительности обслуживания  $b_j$  заявок в соответствующем узле СеМО.

Характеристики всех  $n$  СМО (время ожидания заявок в очереди и пребывания в системе, длина очереди и число заявок в системе, среднее число занятых приборов и т.д.) представляют собой узловые характеристики СеМО.

Среднее время ожидания заявок в очереди может быть рассчитано с использованием выражения (4.8) для многоканальных СМО типа М/М/Ν или выражения (4.1) для одноканальных СМО типа М/М/1, остальные характеристики узла ( $j=1,...,n$ ) с использованием фундаментальных соотношений, представленных в п.3.4.3, а именно:

- нагрузка в узле  $j$ , показывающая среднее число занятых приборов:  
 $y_j = \lambda_j b_j$ ;
- загрузка узла  $j$ :  $\rho_j = \min(y_j / K_j; 1)$ , где  $K_j$  – число обслуживающих приборов в узле  $j$ ;
- коэффициент простоя узла:  $\pi_j = 1 - \rho_j$ ;
- время пребывания заявок в узле:  $u_j = w_j + b_j$ ;
- длина очереди заявок:  $l_j = \lambda_j w_j$ ;
- число заявок в узле (в очереди и на обслуживании в приборе):  
 $m_j = \lambda_j u_j$ .

Рассчитанные таким образом характеристики отдельных СМО в точности соответствуют узловым характеристикам исходной СеМО, то есть в отношении своих характеристик модель массового обслуживания, представляющая собой совокупность независимых СМО (каждая СМО рассматривается независимо от других), строго эквивалентна исходной разомкнутой СеМО в целом.

### Расчет сетевых характеристик РСемо

Сетевые характеристики, описывающие эффективность функционирования СеМО в целом, рассчитываются на основе полученных значений узловых характеристик.

В состав сетевых характеристик входят:

- среднее число заявок, ожидающих обслуживания в сети, и среднее число заявок, находящихся в сети:

$$L = \sum_{j=1}^n l_j ; \quad M = \sum_{j=1}^n m_j ,$$

где  $l_j$  – средняя длина очереди и  $m_j$  – среднее число заявок в узле  $j$ ;

- среднее время ожидания и среднее время пребывания заявок в сети:

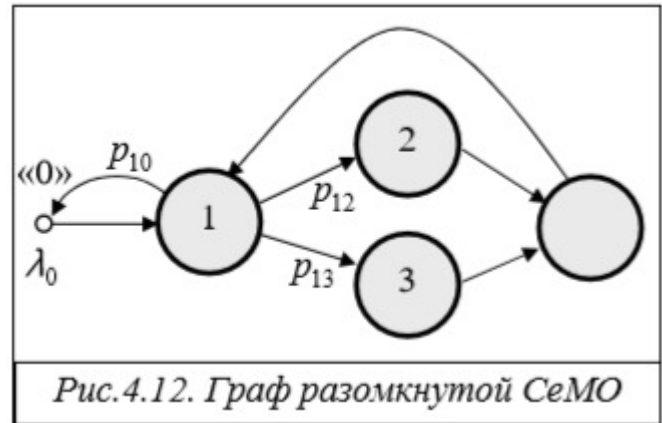
$$W = \sum_{j=1}^n \alpha_j w_j ;$$

$$U = \sum_{j=1}^n \alpha_j u_j ,$$

где  $w_j$  и  $u_j$  – соответственно среднее время ожидания и среднее время пребывания заявок в узле  $j$ ;  $\alpha_j$  – коэффициент передачи для узла  $j$ , показывающий среднее число попаданий заявки в узел  $j$  за время её нахождения в сети.

**Пример 4.2.** Проиллюстрируем изложенный метод расчета характеристик функционирования линейных разомкнутых однородных экспоненциальных СеМО на примере СеМО с четырьмя узлами ( $n = 4$ ), граф которой представлен на рис.4.12. Связи между узлами СеМО описываются следующей матрицей вероятностей передач:

	0	1	2	3	4
0	0	1	0	0	0
1	0,1	0	0,2	0,7	0
2	0	0	0	0	1
3	0	0	0	0	1
4	0	1	0	0	0



В РСемо поступает простейший поток заявок с интенсивностью  $\lambda_0 = 0,1 \text{ с}^{-1}$ . Положим, что все узлы СеМО – одноканальные, а средние длительности обслуживания заявок в узлах соответственно равны:  $b_1=0.8\text{с}$ ;  $b_2=2\text{с}$ ;  $b_3=0.4\text{с}$ ;  $b_4 = 0.3\text{с}$ .

Система линейных алгебраических уравнений для расчёта интенсивностей потоков заявок в узлах СеМО, согласно (4.16), имеет вид:

$$\left. \begin{aligned} \lambda_0 &= p_{10} \lambda_1 = 0,1 \lambda_1 \\ \lambda_1 &= p_{01} \lambda_0 + p_{41} \lambda_4 = \lambda_0 + \lambda_4 \\ \lambda_2 &= p_{12} \lambda_1 = 0,2 \lambda_1 \\ \lambda_3 &= p_{13} \lambda_1 = 0,7 \lambda_1 \\ \lambda_4 &= p_{24} \lambda_2 + p_{34} \lambda_3 = \lambda_2 + \lambda_3 \end{aligned} \right\} ,$$

Решая эту систему уравнений, получим следующие значения интенсивностей:  $\lambda_1 = 1 \text{ с}^{-1}$ ,  $\lambda_2 = 0,2 \text{ с}^{-1}$ ,  $\lambda_3 = 0,7 \text{ с}^{-1}$ ,  $\lambda_4 = 0,9 \text{ с}^{-1}$ . Тогда коэффициенты передач будут равны:  $\alpha_1 = \lambda_1 / \lambda_0 = 10$ ;  $\alpha_2 = \lambda_2 / \lambda_0 = 2$ ;  $\alpha_3 = \lambda_3 / \lambda_0 = 7$ ;  $\alpha_4 = \lambda_4 / \lambda_0 = 9$ .

Определим предельную интенсивность поступления заявок в разомкнутую СеМО, при которой в сети отсутствуют перегрузки. Для этого воспользуемся выражением (3.25), определяющим условие отсутствия перегрузок в РСемо:

$$\lambda_0 < \min \left( \frac{K_1}{\alpha_1 b_1}, \frac{K_2}{\alpha_2 b_2}, \frac{K_3}{\alpha_3 b_3}, \frac{K_4}{\alpha_4 b_4} \right) = 0,125 \text{ с}^{-1}.$$

РСемо работает без перегрузок, поскольку данное условие выполняется.

В соответствии с эквивалентным преобразованием представим рассматриваемую экспоненциальную разомкнутую СеМО в виде 4-х независимых СМО типа М/М/1, в которые поступают простейшие потоки заявок

соответственно с интенсивностями:  $\lambda_1 = 1 \text{ с}^{-1}$ ,  $\lambda_2 = 0,2 \text{ с}^{-1}$ ,  $\lambda_3 = 0,7 \text{ с}^{-1}$ ,  $\lambda_4 = 0,9 \text{ с}^{-1}$ , а средние длительности обслуживания заявок в СМО совпадают с длительностями обслуживания в соответствующих узлах СеМО:  $b_1 = 0,8 \text{ с}$ ;  $b_2 = 2 \text{ с}$ ;  $b_3 = 0,4 \text{ с}$ ;  $b_4 = 0,3 \text{ с}$ .

Значения узловых характеристик СеМО, рассчитанные с использованием выражения (4.1) для среднего времени ожидания заявок в очереди СМО типа М/М/1 и фундаментальных соотношений, представленных в п.3.4.3, приведены в табл.4.1.

Таблица 4.1

Узловые характеристики $u$	Расчётные формулы	Узел 1	Узел 2	Узел 3	Узел 4
Нагрузка	$y_j = \lambda_j b_j$	0,8	0,4	0,28	0,27
Загрузка	$\rho_j = \min(y_j / K_j; 1)$	0,8	0,4	0,28	0,27
Коэф-т простоя	$\pi_j = 1 - \rho_j$	0,2	0,6	0,72	0,73
Время ожидания	$w_j = \rho_j b_j / (1 - \rho_j)$	3,2	1,33	0,16	0,11
Время пребывания	$u_j = w_j + b_j$	4	3,33	0,56	0,41
Длина очереди	$l_j = \lambda_j w_j$	3,2	0,27	0,11	0,10
Число заявок в узле	$m_j = \lambda_j u_j$	4	0,67	0,39	0,37

В табл.4.2 представлены математические зависимости и полученные на их основе значения сетевых характеристик, рассчитанные с учётом найденных значений узловых характеристик.

Таблица 4.2

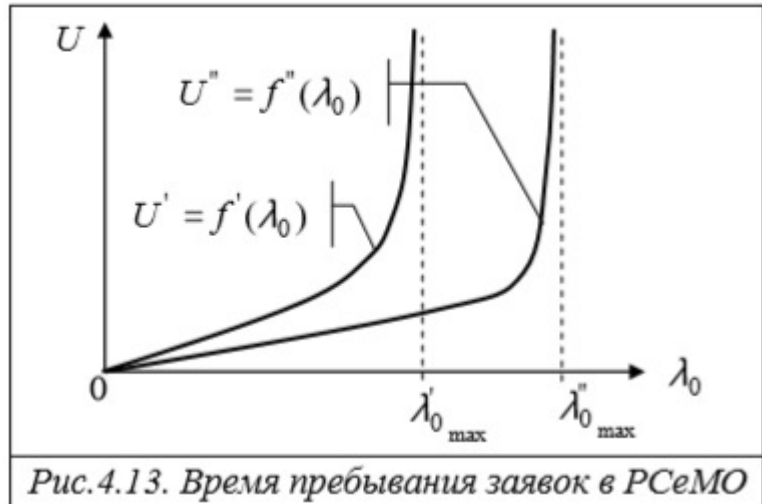
Сетевые характеристики $u$	Расчётные формулы	Значения
Время ожидания в сети	$W = \sum_{j=1}^n \alpha_j w_j$	36,75
Время пребывания в сети	$U = \sum_{j=1}^n \alpha_j u_j$	54,25
Число заявок в состоянии ожидания	$L = \sum_{j=1}^n l_j$	3,68
Число заявок в сети	$M = \sum_{j=1}^n m_j$	5,43



## Анализ свойств разомкнутых СеМО

Свойства разомкнутых СеМО определяются значениями узловых и сетевых характеристик, связанных между собой зависимостями, представленными в разделе 3. Наибольший интерес представляют свойства сети в целом, поскольку свойства отдельных узлов СеМО аналогичны свойствам соответствующих одноканальных и многоканальных СМО.

На рис. 4.13 показана зависимость основной сетевой характеристики РСеМО – среднего времени пребывания  $U$  заявок в сети от интенсивности  $\lambda_0$  поступления заявок в сеть. Зависимость  $U = f(\lambda_0)$  аналогична зависимости среднего времени пребывания заявок в СМО от загрузки системы, изменение которой может быть обусловлено, в частности, изменением интенсивности поступления заявок в СМО. Как и в СМО, имеется некоторое предельное значение интенсивности  $\lambda_{0\max}$  при котором среднее время пребывания заявок в сети становится бесконечно большим, что свидетельствует о перегрузке в СеМО. Выше (см. п. 3.4.2) показано, что в РСеМО отсутствуют перегрузки, если они отсутствуют во всех узлах сети, то есть перегрузка в разомкнутой СеМО наступает в том случае, когда загрузка одного из узлов сети становится равной единице. Такой узел называется «узким местом» и очередь заявок перед ним со временем растёт до бесконечности и, как следствие, становится бесконечным число заявок в разомкнутой СеМО.



Для того чтобы избавиться в РСеМО от перегрузки, необходимо *разгрузить* «узкое место». Это может быть достигнуто следующими способами:

- увеличением скорости работы (быстродействия) обслуживающего прибора;
- увеличением числа обслуживающих приборов в узле.

Любой из этих способов позволяет увеличить производительность СеМО в целом и, как следствие, улучшить характеристики сети. Зависимость среднего времени пребывания  $U$  заявок в сети от интенсивности  $\lambda_0$  поступления заявок в сеть принимает вид  $U'' = f''(\lambda_0)$ , то есть время пребывания заявок при одной и той же интенсивности  $\lambda_0$  становится меньше (поскольку сеть имеет большую производительность), а предельное значение интенсивности  $\lambda_{0\max}''$ , при котором наступает перегрузка СеМО, становится больше:  $\lambda_{0\max}'' > \lambda_{0\max}'$ . При этом появляется новое узкое место в СеМО, и дальнейшее улучшение сети может быть достигнуто путём разгрузки нового узкого места. Очевидно, что если СеМО является моделью реальной технической системы, разгрузка узкого места за счёт увеличения скорости работы обслуживающего прибора или числа приборов в



узле означает увеличение стоимости реальной системы.

Существует ещё один способ разгрузки узкого места СеМО, заключающийся в *уменьшении вероятности передачи* заявок к узлу, являющемуся узким местом. Этот способ часто используется в реальных системах и обычно не связан с увеличением стоимости системы. Например, в вычислительной системе изменение вероятностей передач к накопителям внешней памяти может быть достигнуто за счет перераспределения файлов между накопителями: наиболее часто используемые файлы, расположенные в наиболее загруженном накопителе, переносятся в наименее загруженный накопитель. При этом уменьшается количество обращений к загруженному накопителю (коэффициент передачи соответствующего узла СеМО).

Характер зависимостей других сетевых характеристик (времени ожидания, числа заявок в сети и в состоянии ожидания) разомкнутой СеМО от интенсивности поступления заявок аналогичен показанному на рис. 4.13.

**Пример 4.3.** Проиллюстрируем способы разгрузки узкого места и получаемый от этого эффект для четырёхузловой разомкнутой СеМО, рассмотренной в примере 4.2. Там же было показано, что интенсивность поступления заявок в разомкнутую СеМО, при которой в сети отсутствуют перегрузки, должна удовлетворять условию:  $\lambda_0 < 0,125 \text{ с}^{-1}$ .

1. Рассчитаем сначала характеристики РСеМО, работающей в области загрузок, близких к 1, для чего положим, что интенсивность потока поступающих в сеть заявок равна  $\lambda_0 = 0,12 \text{ с}^{-1}$

Тогда интенсивности потоков заявок в узлы РСеМО соответственно будут равны  $\lambda_1 = \alpha_1 \lambda_0 = 1,2 \text{ с}^{-1}$ ,  $\lambda_2 = \alpha_2 \lambda_0 = 0,24 \text{ с}^{-1}$ ,  $\lambda_3 = \alpha_3 \lambda_0 = 0,84 \text{ с}^{-1}$ ,  $\lambda_4 = \alpha_4 \lambda_0 = 1,08 \text{ с}^{-1}$ , а средние длительности обслуживания заявок, как и ранее, будут равны:  $b_1 = 0,8 \text{ с}$ ;  $b_2 = 2 \text{ с}$ ;  $b_3 = 0,4 \text{ с}$ ;  $b_4 = 0,3 \text{ с}$ .

Рассчитанные значения узловых и сетевых характеристик СеМО приведены в табл.4.3.

Таблица 4.3

Характеристики	Узел 1	Узел 2	Узел 3	Узел 4	СеМО
Нагрузка	0,96	0,48	0,336	0,324	2,10
Загрузка	0,96	0,48	0,336	0,324	2,10
Время ожидания	19,2	1,85	0,202	0,144	198,4
Время пребывания	20	3,85	0,602	0,444	215,9
Длина очереди	23,04	0,44	0,170	0,155	23,8
Число заявок	24	0,92	0,506	0,479	25,9

Анализ представленных результатов показывает, что увеличение интенсивности поступления заявок в РСеМО всего лишь на 20% до значения  $\lambda_0 = 0,12 \text{ с}^{-1}$ , привело к резкому росту значений сетевых характеристик. В частности, среднее время пребывания заявок в сети выросло в 4 раза, а число заявок, находящихся в очередях – почти в 6,5 раз. Это говорит о том, что СеМО работает в области больших загрузок, где незначительное увеличение нагрузки приводит к существенному изменению характеристик обслуживания заявок. Наиболее загруженным узлом СеМО, то есть узким местом, является узел 1, загрузка которого много больше загрузок других узлов и составляет  $\rho_1 = 0,96$ . Именно в этом узле характеристики обслуживания заявок выросли наиболее

существенно: среднее время пребывания заявок в 5 раз (с 4 до 20 секунд), а средняя длина очереди – более чем в 7 раз (с 3,2 до 23 заявок).

2. Для улучшения характеристик обслуживания заявок в РСМО необходимо разгрузить узкое место сети, которым является узел 1. Для этого увеличим скорость работы обслуживающего прибора в 2 раза, что, в конечном счете, приведёт к уменьшению длительности обслуживания заявок в 2 раза, которая станет равной  $b_1 = 0,4$  с.

Рассчитанные значения узловых и сетевых характеристик СеМО после разгрузки узкого места приведены в табл.4.4.

Анализ представленных результатов показывает, что разгрузка узкого места позволила существенно уменьшить значения сетевых характеристик: среднее время пребывания заявок в сети уменьшилось более чем в 9 раз, а число заявок, находящихся в очередях – почти в 20 раз. Отметим, что изменение длительности обслуживания заявок в узле 1 привело к изменению узловых характеристик только этого узла; узловые характеристики остальных узлов не изменились. Это является следствием независимого функционирования узлов экспоненциальной разомкнутой СеМО, что фактически и позволяет использовать метод расчёта характеристик сети, основанный на декомпозиции, то есть представлении сети в виде совокупности независимых СМО.

Таблица 4.4

Узловые характеристики $u$	Узел 1	Узел 2	Узел 3	Узел 4	СеМО
Нагрузка	0,48	0,48	0,336	0,324	1,62
Загрузка	0,48	0,48	0,336	0,324	1,62
Время ожидания	0,369	1,846	0,202	0,144	10,1
Время пребывания	0,769	3,846	0,602	0,444	23,6
Длина очереди	0,443	0,443	0,170	0,155	1,21
Число заявок	0,923	0,923	0,506	0,479	2,83

3. Для сравнения выполним разгрузку узкого места другим способом, а именно: увеличим число обслуживающих приборов в узле 1 с одного до двух:  $K_1 = 2$ , сохранив прежнее значение длительности обслуживания одним прибором:  $b_1 = 0,8$  с.

Рассчитанные значения узловых и сетевых характеристик СеМО после разгрузки узкого места приведены в табл.4.5.

Таблица 4.5

Узловые характеристики $u$	Узел 1	Узел 2	Узел 3	Узел 4	СеМО
Нагрузка	0,96	0,48	0,336	0,324	2,10
Загрузка	0,48	0,48	0,336	0,324	1,62
Время ожидания	0,288	1,846	0,202	0,144	9,28
Время пребывания	1,088	3,846	0,602	0,444	26,78
Длина очереди	0,346	0,443	0,170	0,155	1,11
Число заявок	1,306	0,923	0,506	0,479	3,21

Сравним полученные значения сетевых характеристик со значениями, представленными в табл. 4.4 для первого способа разгрузки узкого места за счёт уменьшения длительности обслуживания заявок. При втором способе разгрузки узкого места за счёт увеличения числа обслуживающих приборов ( $K_1=2$ ;  $b_1 = 0,8$  с) среднее время ожидания заявок в сети несколько уменьшилось по сравнению с первым способом ( $K_1 = 1$ ;  $b_1 = 0,4$  с). В то же время, среднее время пребывания заявок в РСеМО увеличились более чем на 10%, что обусловлено большей длительностью обслуживания заявок ( $b_1 = 0,8$  с) в каждом из приборов двухканального узла 1 по сравнению с одноканальным узлом при первом способе ( $b_1 = 0,4$  с). Как и в предыдущем случае, изменение числа обслуживающих приборов в узле 1 привело к изменению узловых характеристик только этого узла.

Конец лекций 2019,2020,2021 г

### 3.2. Замкнутые экспоненциальные СеМО с однородным потоком заявок

#### Описание замкнутых СеМО

Рассмотрим замкнутую экспоненциальную сеть массового обслуживания с однородным потоком заявок при следующих предположениях:

- 1) замкнутая СеМО (ЗСеМО) произвольной топологии содержит  $n$  узлов;
- 2) после завершения обслуживания в каком-либо узле передача заявки в другой узел происходит мгновенно;
- 3) все узлы замкнутой СеМО одноканальные;
- 4) в СеМО циркулирует постоянное число заявок;
- 5) длительности обслуживания заявок во всех узлах сети представляют собой случайные величины, распределенные по экспоненциальному закону;
- 6) ёмкость накопителя в каждом узле СеМО достаточна для хранения всех заявок, циркулирующих в сети, что означает отсутствие отказов поступающим заявкам при их постановке в очередь любого узла (в частности, можно считать, что ёмкость накопителя в каждом узле равна числу заявок, циркулирующих в сети);
- 7) обслуживающий прибор любого узла не простаивает, если в его накопителе имеется хотя бы одна заявка, причем после завершения обслуживания очередной заявки мгновенно из накопителя выбирается следующая заявка;
- 8) в каждом узле сети заявки из накопителя выбираются в соответствии с беспriorитетной дисциплиной обслуживания в порядке поступления (ОПП) по правилу «первым пришел – первым обслужен» (FIFO – First In First Out).

Для описания линейных замкнутых однородных экспоненциальных СеМО необходимо задать такую же совокупность параметров, как и для разомкнутых СеМО, с единственным отличием, заключающимся в том, что вместо интенсивности источника заявок следует задать число заявок, циркулирующих в ЗСеМО. Таким образом, совокупность параметров для замкнутых СеМО будет иметь следующий вид:

- число узлов в сети:  $n$ ;

- число обслуживающих приборов в узлах сети:  $K_1, \dots, K_n$ ;
  - матрица вероятностей  $P = [p_{ij} \mid i, j = 0, 1, \dots, n]$ , где  $p_{ij}$  – вероятность передачи заявки из узла  $i$  в узел  $j$ ;
  - число заявок  $M$ , циркулирующих в ЗСеМО;
  - средние длительности обслуживания заявок в узлах сети:  $b_1, \dots, b_n$ .
- На основе перечисленных параметров могут быть рассчитаны узловые и сетевые характеристики, описывающие эффективность функционирования соответственно узлов и ЗСеМО в целом.

Расчёт характеристик функционирования линейных замкнутых однородных экспоненциальных СеМО с одноканальными узлами базируется на так называемой «теореме о прибытии» и проводится с использованием метода средних значений в два этапа:

- расчет коэффициентов передач в узлах замкнутой СеМО;
- расчет характеристик ЗСеМО.

### Расчет коэффициентов передач в узлах ЗСеМО

Для замкнутой СеМО на первом этапе рассчитываются только коэффициенты передач. Интенсивности потоков заявок в узлах ЗСеМО не могут быть рассчитаны, как в РСеМО, поскольку для ЗСеМО изначально не известна интенсивность  $\lambda_0$ , которая является не параметром, задаваемым в составе исходных данных, а характеристикой, представляющей собой производительность ЗСеМО и определяемой в процессе анализа эффективности функционирования ЗСеМО.

Для расчёта коэффициентов передач  $\alpha_1, \dots, \alpha_n$  после некоторых преобразований можно воспользоваться той же системой линейных алгебраических уравнений (4.16). Для этого в левой и правой части выражения (4.16) представим интенсивности в виде  $\lambda_j = \alpha_j \lambda_0$ . Разделив левую и правую часть выражения (4.16) на  $\lambda_0$ , окончательно получим систему линейных алгебраических уравнений относительно  $\alpha_1, \dots, \alpha_n$ :

$$\alpha_j = \sum_{i=0}^n p_{ij} \alpha_j \quad (i = 0, 1, \dots, n). \quad (4.17)$$

Полагая  $\alpha_0 = 1$ , можно найти корни системы уравнений, численно определяющие значения  $\alpha_1, \dots, \alpha_n$ .

### Расчет характеристик ЗСеМО

Характеристики ЗСеМО могут быть рассчитаны с использованием марковских процессов, поскольку количество состояний марковского процесса, в отличие от РСеМО, не бесконечно и равно числу сочетаний  $C_{M+n-1}^M$  где  $n$  – число узлов в ЗСеМО и  $M$  – число заявок, циркулирующих в ЗСеМО. При этом основная трудность заключается в определении вероятностей состояний сети  $P(M_1, \dots, M_n)$  в случае большой ее размерности ( $n > 5$ ;  $M > 5$ ), когда число состояний оказывается значительным. При выполнении расчетов на ЭВМ это, во многих случаях, приводит к потере значимости в процессе промежуточных вычислений и, следовательно, к невозможности получения конечных результатов.

От указанного недостатка свободен метод средних значений, позволяющий вычислять средние характеристики функционирования экспоненциальных СеМО

на основе сравнительно простых рекуррентных соотношений.

Положим, что замкнутая однородная СеМО содержит  $n$  **одноканальных** узлов, длительности обслуживания заявок в которых распределены по экспоненциальному закону со средними значениями  $b_1, \dots, b_n$  соответственно. Пусть для каждого узла  $i$  сети известно среднее число попаданий заявки в данный узел за время ее нахождения в сети, то есть коэффициент передачи  $\alpha_i$ , который, если конфигурация сети задана матрицей вероятностей передач

$\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$ , определяется в результате решения системы линейных алгебраических уравнений (4.17).

Обозначим:  $u_i$  – среднее время пребывания заявки в узле  $i$  за время пребывания в сети;  $m_i$  – среднее число заявок в узле  $i$  ( $i = 1, \dots, n$ );  $\lambda_0$  – производительность замкнутой сети. Очевидно, что эти величины зависят от числа заявок  $M$ , циркулирующих в замкнутой сети, то есть  $u_i = u_i(M)$ ;  $m_i = m_i(M)$ ;  $\lambda_0 = \lambda_0(M)$ .

Можно показать, что имеют место следующие соотношения:

$$u_i(M) = b_i [1 + m_i(M - 1)]; \quad (4.18)$$

$$U(M) = \sum_i \alpha_i u_i(M); \quad (4.19)$$

$$\lambda_0(M) = \frac{M}{U(M)}; \quad (4.20)$$

$$m_i(M) = \alpha_i \lambda_0(M) u_i(M), \quad (4.21)$$

где  $U(M)$  – среднее время пребывания заявок в сети при условии нахождения в ней  $M$  заявок;  $m_i(0) = 0$ .

Выражение (4.18) получено на основе так называемой *теоремы о прибытии* [1], утверждающей, что в замкнутой экспоненциальной сети с одноканальными узлами, в которой циркулируют  $M$  заявок, стационарная вероятность состояния любого узла в момент поступления в него новой заявки совпадает со стационарной вероятностью того же состояния рассматриваемого узла в сети, в которой циркулирует на одну заявку меньше, то есть  $(M-1)$  заявок. Это означает, что в сети с  $M$  заявками среднее число ЗАЯВОК  $m_i(M)$ , находящихся в узле  $i$  в момент поступления в этот узел новой заявки, равно

$m_i(M - 1)$ . Тогда среднее время пребывания в узле  $i$  поступившей заявки будет складываться из среднего времени обслуживания всех ранее поступивших и находящихся в узле  $i$  заявок и средней длительности обслуживания рассматриваемой заявки:

$$u_i(M) = b_i m_i(M - 1) + b_i = b_i [1 + m_i(M - 1)].$$

В этом выражении учтено, что среднее время дообслуживания заявки, находящейся в приборе на момент поступления рассматриваемой заявки, равно средней длительности обслуживания  $b_i$  в силу свойства отсутствия последствия, присущего экспоненциальному закону. Среднее время пребывания заявки в узле  $i$  за время ее нахождения в сети, учитывающее число попаданий  $\alpha_i$  заявки в данный узел, равно

$$U_i(M) = \alpha_i u_i(M).$$

Выражения (4.19) и (4.20) представляют собой формулы Литтла для сети, а выражение (4.21) – для узла  $i$ , где  $\lambda_i(M) = \alpha_i \lambda_0(M)$  – интенсивность потока заявок

в узел  $i$  ( $i = 1, \dots, n$ ).

На основе рекуррентных соотношений (4.18) – (4.21) последовательно для  $M = 1, 2, \dots, M^*$ , где  $M^*$  – заданное число заявок в замкнутой сети, могут быть рассчитаны средние значения характеристик замкнутой экспоненциальной СеМО.

Заметим, что приведенный метод расчета является *точным* для замкнутых экспоненциальных СеМО с *одноканальными* узлами.

**Пример 4.4.** Рассчитаем характеристики замкнутой однородной экспоненциальной СеМО, полученной путём преобразования разомкнутой СеМО (рис. 4.12), рассмотренной в Примере 4.2, в замкнутую. Положим, что «нулевая точка», отображающая завершение обслуживания заявок в сети и мгновенное формирование новой заявки, выбрана на дуге, выходящей из узла 1 и входящей снова в этот же узел (рис.4.14). Напомним, что в ЗСеМО относительно «нулевой точки» рассчитываются временные сетевые характеристики: время нахождения в состоянии ожидания и время пребывания заявок в сети, а также производительность ЗСеМО.

ЗСеМО содержит  $n = 4$  одноканальных узла, связи между которыми описываются той же матрицей вероятностей передач:

		0	1	2	3	4
	0	0	1	0	0	0
	1	0,1	0	0,2	0,7	0
	2	0	0	0	0	1
<b>P =</b>	3	0	0	0	0	1
	4	0	1	0	0	0

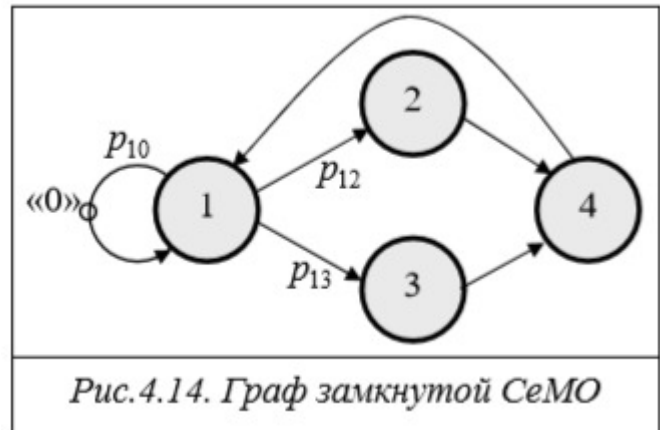


Рис.4.14. Граф замкнутой СеМО

Следовательно, коэффициенты передач для всех узлов, рассчитываемые путём решения системы линейных алгебраических уравнений (4.17), будут иметь те же самые значения:  $\alpha_1 = 10$ ;  $\alpha_2 = 2$ ;  $\alpha_3 = 7$ ;  $\alpha_4 = 9$ .

В ЗСеМО циркулирует  $M$  заявок, средние длительности обслуживания которых в узлах равны:  $b_1 = 0,8$  с;  $b_2 = 2$  с;  $b_3 = 0,4$  с;  $b_4 = 0,3$  с.

Ниже в табл.4.6 представлены значения времени пребывания  $u_i(M)$  и числа заявок  $m_i(M)$  в узлах сети, а также среднего времени пребывания  $U(M)$  заявок в сети и производительности  $\lambda_0(M)$ , рассчитанные на основе выражений (4.18) – (4.21), для числа циркулирующих в сети заявок  $M = 1, 2, \dots, 6$ . Корректность выполненных расчетов подтверждается тем, что для всех  $M = 1, 2, \dots, 6$  выполняется проверочное условие:

$$\sum_{i=1}^4 m_i(M) = M.$$

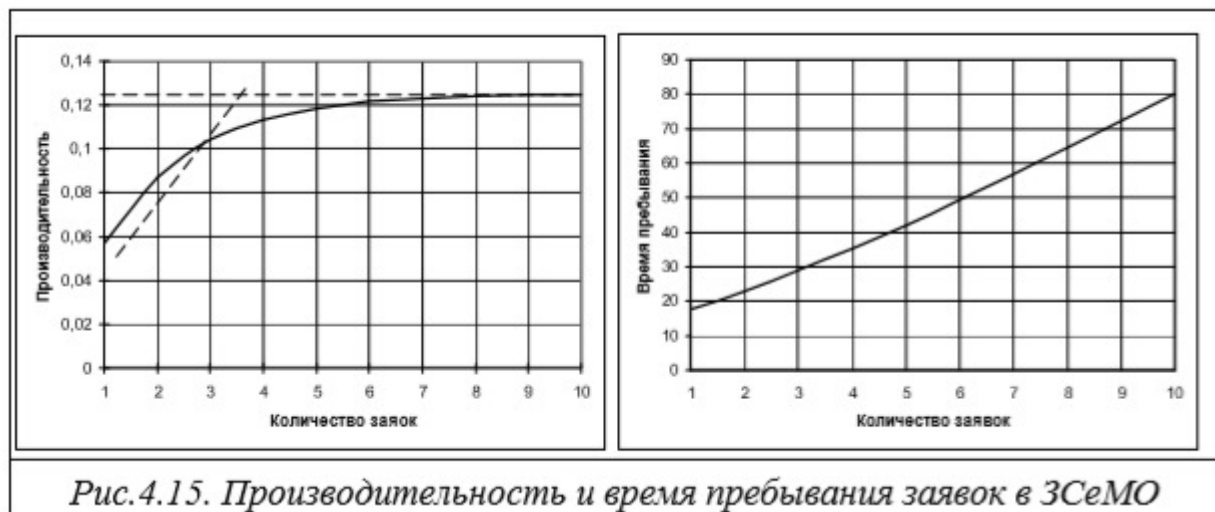
Таблица 4.6

$M$	$i$	$u_i(M)$	$U(M)$	$\lambda_0(M)$	$m_i(M)$
1	1	0,8	17,5	0,057	0,46
	2	2,0			0,23
	3	0,4			0,16
	4	0,3			0,15

2	1	1,17	22,94	0,087	1,02
	2	2,46			0,43
	3	0,46			0,28
	4	0,35			0,27
3	1	1,61	28,87	0,104	1,68
	2	2,86			0,59
	3	0,51			0,37
	4	0,38			0,36
4	1	2,14	35,29	0,113	2,43
	2	3,19			0,72
	3	0,55			0,44
	4	0,41			0,42
5	1	2,74	42,14	0,119	3,25
	2	3,45			0,82
	3	0,57			0,48
	4	0,42			0,45
6	1	3,40	49,35	0,122	4,14
	2	3,63			0,88
	3	0,59			0,50
	4	0,44			0,48

На рис.4.15 представлены зависимости производительности рассматриваемой замкнутой СеМО и среднего времени пребывания заявок в сети от количества

$M = 1, 10$  циркулирующих заявок. Анализ полученных результатов показывает, что все характеристики, включая производительность  $\lambda_0$ , растут с увеличением  $M$ . Производительность сети асимптотически приближается к максимально возможной производительности (пропускной способности ЗСеМО), совпадающей с предельно допустимой интенсивностью поступления заявок в аналогичной разомкнутой СеМО (см. Пример 4.1), при которой в сети отсутствуют перегрузки, и равна  $\lambda_0 = 0,125 \text{ с}^{-1}$ .



Среднее время пребывания заявок в ЗСеМО растёт неограниченно с увеличением количества заявок с сети.

Остальные характеристики замкнутой СеМО (загрузки и коэффициенты

простоя узлов, время ожидания, длины очередей и число заявок в узлах сети, полное время ожидания в сети) могут быть рассчитаны с использованием фундаментальных соотношений, представленных в разделе 3 (п.3.4.3).

### Анализ свойств замкнутых СеМО

Для замкнутых СеМО, как и для разомкнутых, наибольший интерес представляют свойства сети в целом, в частности, влияние циркулирующих в ЗСеМО числа заявок, на такие сетевые характеристики как производительность  $\lambda_0$  замкнутой СеМО и среднее время пребывания  $U$  заявок в сети.

Анализ представленных на рис.4.16, зависимостей позволяет сформулировать следующие выводы.

1. Зависимость  $\lambda_0 = f(M)$  производительности ЗСеМО  $\lambda_0$  от числа  $M$  циркулирующих заявок вначале растёт с увеличением  $M$  до некоторого значения  $M_0$ , после которого рост производительности замедляется, а с дальнейшим увеличением  $M$  производительность сети асимптотически стремится к некоторому предельному значению  $\hat{\lambda}_0$ , представляющему собой *пропускную способность* ЗСеМО. Для объяснения этой зависимости вспомним, что производительность замкнутой сети

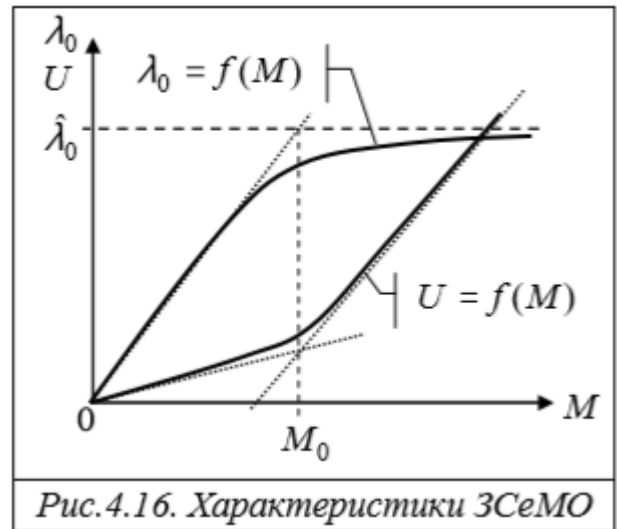


Рис.4.16. Характеристики ЗСеМО

измеряется как интенсивность потока заявок, проходящих через некоторую условную точку, обозначаемую как «0» и расположенную на одной из дуг СеМО, отображающей завершение заявок в сети и мгновенное формирование новой заявки, поступающей в сеть. Выше (см. пример 4.4) было показано, что увеличение числа заявок в замкнутой СеМО приводит к увеличению значений всех сетевых характеристик, включая производительность  $\lambda_0$ . В свою очередь, увеличение производительности приводит к увеличению загрузок узлов СеМО, связанных с интенсивностью  $\lambda_0$  зависимостью:

$$\rho_j = \frac{\alpha_j \lambda_0 b_j}{K_j},$$

где  $\alpha_j, b_j$  и  $K_j$  — соответственно коэффициент передачи, средняя длительность обслуживания и количество приборов в узле  $j = \overline{1, n}$ .

Когда число заявок в ЗСеМО достигает некоторого значения  $M_0$ , нагрузка одного из узлов становится близкой к 1, при этом практически прекращается рост производительности, которая при  $M \rightarrow \infty$  достигает своего предельного значения — пропускной способности  $\hat{\lambda}_0$ . Такой узел представляет собой «узкое место» сети, и значение пропускной способности  $\hat{\lambda}_0$  определяется пропускной способностью узкого места из условия, что нагрузка  $\rho_y$  этого узла равна 1:

$$\rho_y = \frac{\alpha_y \lambda_0 b_y}{K_y} = 1.$$



Отсюда пропускная способность замкнутой СеМО:

$$\hat{\lambda}_0 = \frac{K_y}{\alpha_y b_y},$$

где  $\alpha_y, b_y$  и  $K_y$  - соответственно коэффициент передачи, средняя длительность обслуживания и количество обслуживающих приборов в узле, являющимся узким местом.

Правая часть последнего выражения представляет собой пропускную способность узла, являющегося узким местом сети:  $\mu_y = \frac{K_y}{\alpha_y b_y}$ .

Действительно,  $\alpha_y b_y$  представляет собой полное время обслуживания одной заявки в данном узле с учётом того, что заявка за время нахождения в сети в среднем  $\alpha_y$  раз побывает в данном узле. Тогда величина, обратная  $\alpha_y b_y$ , представляет собой интенсивность обслуживания заявок одним прибором в данном узле:  $\mu_1 = 1/\alpha_y b_y$ , а  $\mu_y = K_y \mu_1$  - интенсивность обслуживания заявок узлом, то есть всеми приборами.

Этот же результат можно получить следующими рассуждениями. Если нагрузка некоторого узла, являющегося узким местом СеМО, становится равной 1, то это означает, что все приборы данного узла постоянно обслуживают заявки, то есть не простаивают. Тогда интенсивность выходящего из этого узла потока заявок будет равна интенсивности обслуживания:  $\lambda_y = \alpha_y \lambda_0$ . Отсюда вытекает, что

производительность ЗСеМО равна  $\lambda_0 = \frac{\lambda_y}{\alpha_y} = \frac{K_y \mu_1}{\alpha_y} = \frac{K_y}{\alpha_y b_y}$ .

2. Среднее время пребывания заявок (рис.4.16) в замкнутой СеМО, как и производительность, растёт с увеличением числа  $M$  циркулирующих в сети заявок, причём вначале наблюдается незначительный рост, а затем, после значения  $0 \text{ ММ} =$ , наблюдается линейный рост времени пребывания.

**a.** Действительно, если в сети циркулирует только одна заявка, то в такой сети не может быть очередей, и время пребывания заявок в СеМО складывается только из времён обслуживания заявок в узлах с учётом коэффициентов передач:

$$b. \quad U = \sum_{i=1}^n \alpha_i b_i .$$

**с.** С увеличением числа заявок  $M$  в узлах ЗСеМО появляются очереди, причём очевидно, что чем больше заявок в сети, тем более длинные очереди образуются в узлах и тем больше время ожидания, а, следовательно, и время пребывания заявок в ЗСеМО.

**d.** Сопоставляя зависимости производительности и среднего времени пребывания заявок от их числа в ЗСеМО, можно сделать следующий вывод: увеличение числа заявок в сети, с одной стороны, приводит к увеличению производительности, что может рассматриваться как положительный фактор, а, с другой стороны, – к увеличению времени пребывания заявок в сети, что является нежелательным фактором.

**e.** Точка  $M=M_0$  характеризует некоторое граничное значение числа заявок в ЗСеМО. Дальнейшее увеличение числа заявок в сети оказывается нецелесообразным, поскольку приводит к резкому увеличению времени пребывания заявок в ЗСеМО при незначительном увеличении производительности сети.

3. Когда загрузка узкого места становится равной единице, дальнейший рост производительности за счёт увеличения числа заявок в ЗСеМО невозможен. Для увеличения производительности ЗСеМО, как и в РСеМО, необходимо разгрузить узкое место, то есть уменьшить загрузку:

$\rho_y = \frac{\alpha_y \lambda_0 b_y}{K_y} = 1$ , что при одной и той же производительности может быть достигнуто:

- уменьшением длительности обслуживания заявок  $b_y$ , например за счет увеличения скорости работы (быстродействия) обслуживающего прибора;
- увеличением числа обслуживающих приборов  $K_y$  в узле;
- уменьшением коэффициента передачи  $\alpha_y$  или, что то же самое, вероятности передачи заявок к узлу, являющемуся узким местом.

Если до разгрузки узкого места зависимость производительности ЗСеМО от числа заявок в сети имела вид  $\lambda'_0 = f(M)$  (рис 4.17), а пропускная способность была равна  $\hat{\lambda}'_0$ , то после разгрузки зависимость производительности от числа заявок будет иметь вид  $\lambda''_0 = f(M)$ , а пропускная способность станет равной  $\hat{\lambda}''_0 > \hat{\lambda}'_0$ . При этом граничное значение числа заявок в ЗСеМО увеличится:  $M''_0 > M'_0$ .

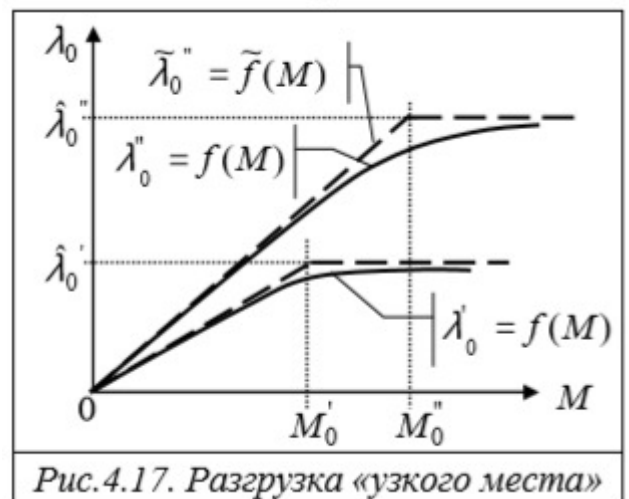


Рис.4.17. Разгрузка «узкого места»

Следует отметить, что к рассматриваемой зависимости производительности ЗСеМО  $\lambda_0$  от числа  $M$  циркулирующих в сети заявок может быть применена линейная аппроксимация  $\tilde{\lambda}_0'' = \tilde{f}(M)$ , показанная на рис.4.17 в виде пунктирных линий и представляющая собой верхнюю границу производительности ЗСеМО. Последнее означает, что производительность ЗСеМО будет не больше, чем рассчитанное верхнее значение.

Нетрудно представить себе и изобразить на графике, как изменится зависимость среднего времени пребывания заявок в замкнутой СеМО от числа циркулирующих в сети заявок после разгрузки узкого места.

Отметим, что в некоторых случаях разгрузка узкого места не приводит к улучшению характеристик СеМО, в частности, к увеличению производительности. Обычно это связано с тем, что в СеМО может существовать несколько узлов, являющихся «узкими местами». Условием этого является равенство загрузок узлов:  $\rho_i = \rho_j$  или  $\frac{\alpha_i \lambda_0 b_i}{K_i} = \frac{\alpha_j \lambda_0 b_j}{K_j}$ , откуда окончательно

получим:  $\frac{\alpha_i b_i}{K_i} = \frac{\alpha_j b_j}{K_j} (i \neq j)$ . В этом случае для улучшения характеристик ЗСеМО необходимо одновременно разгрузить все «узкие места».

Последовательно разгружая узкие места СеМО, мы можем прийти к некоторой «идеальной» сети, в которой загрузки всех узлов одинаковы.

СеМО, в которой загрузки всех узлов равны, называется **сбалансированной**. Сбалансированная СеМО обладает наилучшими характеристиками по сравнению с несбалансированной.

При построении реальных систем, моделями которых служат СеМО, необходимо, по-возможности, строить сбалансированные системы, хотя на практике по многим причинам достичь этого не удаётся.

#### f. Марковские модели сетей массового обслуживания

В данном параграфе подробно рассматриваются марковские модели сетей массового обслуживания (СеМО) с однородным потоком заявок. В качестве примеров представлены разомкнутые и замкнутые экспоненциальные СеМО с накопителями ограниченной ёмкости, а также замкнутые неэкспоненциальные СеМО, в которых длительность обслуживания заявок в одном из узлов распределена по закону Эрланга с коэффициентом вариации  $v < 1$  и гиперэкспоненциальному закону с коэффициентом вариации  $v > 1$ .

Можно показать, что случайный процесс, протекающий в экспоненциальных разомкнутых и замкнутых СеМО при сформулированных предположениях и допущениях является марковским.

Случайный процесс, протекающий в замкнутой неэкспоненциальной сети, не является марковским. Для описания процесса функционирования такой системы в терминах марковских случайных процессов в некоторых случаях можно воспользоваться методом вложенных цепей Маркова, суть которого заключается в том, что функционирование системы рассматривается в определенные моменты времени, образующие цепь Маркова.

Как и для СМО, в каждом примере приводится *описание* исследуемой СеМО и принятые при построении математической модели *предположения и*

допущения, необходимые для того, чтобы протекающий в системе случайный процесс мог быть сведён к марковскому. Разработка Марковской модели включает в себя этапы кодирования состояний случайного процесса, построения размеченного графа переходов, формирования матрицы интенсивностей переходов и системы линейных алгебраических уравнений для расчёта стационарных вероятностей состояний марковского процесса, на основе которых строятся математические зависимости, позволяющие рассчитать наиболее важные характеристики функционирования исследуемых СеМО.

Применение марковских случайных процессов для расчёта характеристик функционирования и исследования свойств СеМО оказывается наиболее результативным:

- для разомкнутых СеМО с накопителями ограниченной ёмкости, в которых заявки теряются при заполненных накопителях;
- для неэкспоненциальных разомкнутых и замкнутых СеМО, в которых длительности обслуживания заявок в узлах распределены по гипоекспоненциальному или гиперэкспоненциальному закону.

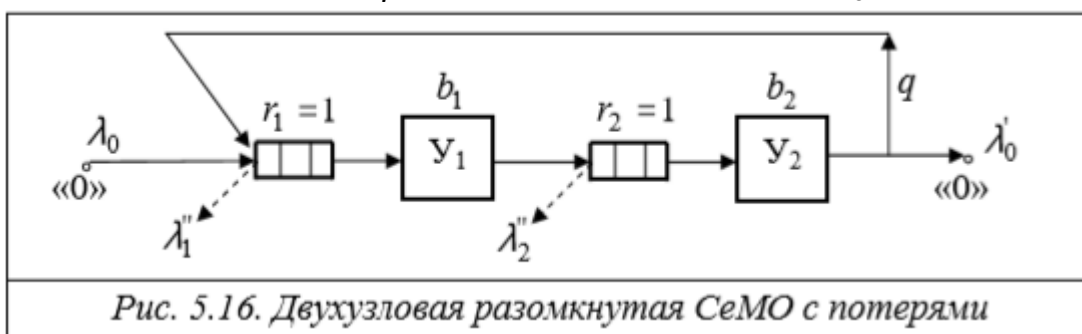
### Разомкнутая экспоненциальная СеМО с накопителями ограниченной ёмкости

Рассмотрим разомкнутую экспоненциальную СеМО с двумя одноканальными узлами, в которую из внешней среды поступает простейший поток заявок с интенсивностью  $\lambda_0$  (рис.5.16). Накопители в обоих узлах имеют ограниченную ёмкость, равную единице:  $r_1 = r_2 = 1$ . Заявка, поступившая в узел и заставшая накопитель заполненным, теряется. Длительности обслуживания в узлах распределены по экспоненциальному закону со средними значениями  $b_1$  и  $b_2$  соответственно. Заявки после обслуживания в узле 2 вероятностью  $q$  направляются в узел 1 и вероятностью  $(1 - q)$  покидают СеМО.

Отметим, что, поскольку заявки в сети могут теряться, рассматриваемая разомкнутая СеМО является *нелинейной*, то есть интенсивности потоков заявок, поступающих в узлы СеМО, не связаны между собой линейной зависимостью (3.5) и, следовательно, не могут быть рассчитаны путём решения системы линейных алгебраических уравнений (4.16).

#### Описание СеМО (рис.5.16).

- Сеть массового обслуживания – разомкнутая двухузловая.
- Узлы 1 и 2 – одноканальные:  $K_1 = K_2 = 1$ .
- Накопители в узлах ограниченной ёмкости:  $r_1 = r_2 = 1$ .
- Дисциплины буферизации в узлах – с потерями заявок, если накопители заполнены.
- Поток заявок однородный с интенсивностью  $\lambda_0$ .



□ Предположения и допущения.

- Поступающие в разомкнутую СеМО заявки образуют *простейший* поток с интенсивностью  $\lambda_0$ .
- Длительности обслуживания заявок в узлах СеМО распределены по *экспоненциальному* закону с параметрами, представляющими собой интенсивности обслуживания:  $\mu_1 = 1/b_1$  и  $\mu_2 = 1/b_2$ .

В разомкнутой СеМО при любой нагрузке существует стационарный режим, так как в узлах сети не может быть бесконечных очередей.

### Кодирование состояний случайного процесса.

Для описания состояний марковского случайного процесса будем использовать распределение заявок между узлами. Закодируем состояния следующим образом:  $(M_1, M_2)$ , где  $M_i = \{0, 1, 2\}$  – количество заявок в узле  $i$  («0» – узел свободен; «1» – на обслуживании в узле находится одна заявка; «2» – в узле находятся две заявки – одна на обслуживания и вторая в накопителе).

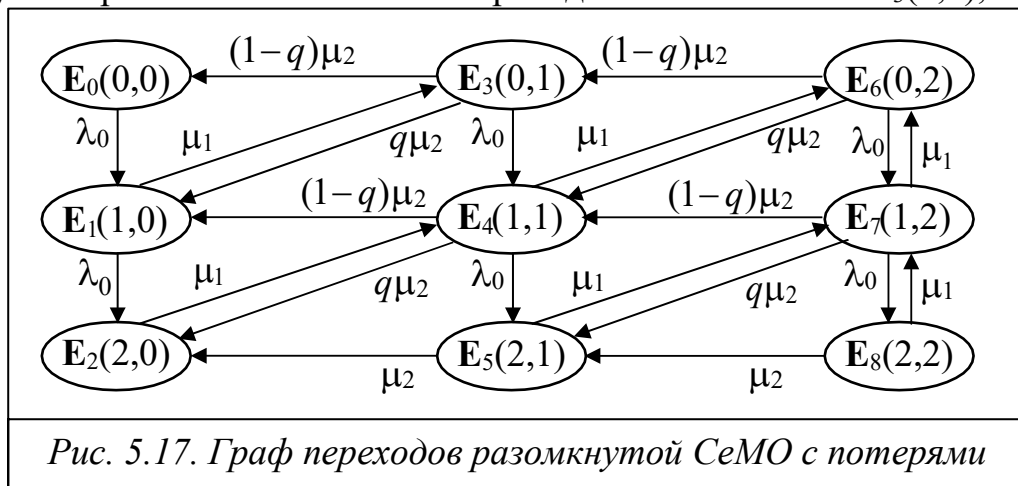
При выбранном способе кодирования система может находиться в следующих состояниях:

- $E_0: (0,0)$  – в СеМО нет ни одной заявки;  $E_1: (1,0)$  – в узле 1 находится одна заявка;  $E_2: (2,0)$  – в узле 1 находятся две заявки;  $E_3: (0,1)$  – в узле 2 находится одна заявка;
- $E_4: (1,1)$  – в узле 1 и 2 находится по одной заявке;
- $E_5: (2,1)$  – две заявки находятся в узле 1 и одна – в узле 2;
- $E_6: (0,2)$  – в узле 2 находятся две заявки;
- $E_7: (1,2)$  – две заявки находятся в узле 2 и одна – в узле 1;
- $E_8: (2,2)$  – в узле 1 и 2 находятся по две заявки.

□ Размеченный граф переходов случайного процесса (рис.5.17).

Построим граф переходов, полагая, что в каждый момент времени может произойти только одно событие (поступление заявки в СеМО или завершение обслуживания заявки в одном из узлов), поскольку вероятность появления двух и более событий в один и тот же момент времени равна нулю.

Следует обратить внимание на переходы из состояний  $E_3(0,1)$ ,  $E_4(1,1)$ ,



$E_6(0,2)$  и  $E_7(1,2)$ , обусловленные завершением обслуживания заявки в узле 2 с интенсивностью  $\mu_2$ . В этих случаях с вероятностью  $q$  заявка может вернуться в

узел 1 и с вероятностью  $(1-q)$  – покинуть СеМО, тогда интенсивности соответствующих переходов будут равны  $q\mu_2$  и  $(1-q)\mu_2$ . Если же случайный процесс находится в состояниях  $E_5(2,1)$  и  $E_8(2,2)$ , то завершение обслуживания заявки в узле 2 приводит к переходу соответственно в состояния  $E_2(2,0)$  и  $E_5(2,1)$  с интенсивностью  $\mu_2$ , что соответствует выходу заявки из СеМО с вероятностью  $(1-q)$  и потере заявки, которая с вероятностью  $q$  будет направлена в узел 1, поскольку в последнем нет места в накопителе. Аналогично, если случайный процесс находится в состояниях  $E_7(1,2)$  и  $E_8(2,2)$ , то завершение обслуживания заявки в узле 1 приводит к переходу соответственно в состояния  $E_6(0,2)$  и  $E_7(1,2)$  с интенсивностью  $\mu_1$ , что соответствует потере заявки, поскольку накопитель узла 1 заполнен.

#### □ Расчет характеристик СеМО.

Не составляя матрицу интенсивностей переходов и не выписывая систему уравнений для определения стационарных вероятностей, получим математические выражения для определения узловых и сетевых характеристик разомкнутой СеМО с потерями при известных значениях стационарных вероятностей состояний  $p_i$  ( $i = 0, 1, \dots, 8$ ).

Заметим, что СеМО с потерями относится к классу *нелинейных* сетевых моделей, расчёт характеристик которых связан с определёнными проблемами, в частности, с необходимостью детального анализа потоков заявок и с невозможностью применения в ряде случаев фундаментальных соотношений для расчёта сетевых характеристик. Кроме того, процесс формирования математических зависимостей для каждой конкретной нелинейной СеМО может существенно отличаться.

В связи с этим, ниже достаточно подробно рассматривается процесс получения математических выражений для расчёта узловых и сетевых характеристик нелинейной разомкнутой СеМО, представленной на рис.5.16.

*Узловые характеристики* СеМО рассчитываются в такой последовательности:

1) загрузки узлов определяются как суммы вероятностей состояний, в которых соответствующий узел занят обслуживанием заявок:

$$\rho_1 = p_1 + p_2 + p_4 + p_5 + p_7 + p_8; \rho_2 = p_3 + p_4 + p_5 + p_6 + p_7 + p_8;$$

2) коэффициенты простоя узлов:  $\eta_1 = 1 - \rho_1; \eta_2 = 1 - \rho_2$ ;

3) среднее число заявок в очередях:

$$l_1 = p_2 + p_5 + p_8; l_2 = p_6 + p_7 + p_8;$$

4) среднее число заявок в узлах:

$$m_1 = p_1 + p_4 + p_7 + 2(p_2 + p_5 + p_8) = l_1 + \rho_1;$$

$$m_2 = p_3 + p_4 + p_5 + 2(p_6 + p_7 + p_8) = l_2 + \rho_2;$$

5) производительности узлов (интенсивность обслуженных заявок на выходе узлов):  $\lambda'_1 = \frac{\rho_1}{b_1} = \rho_1 \mu_1; \quad \lambda'_2 = \frac{\rho_2}{b_2} = \rho_2 \mu_2;$

6) вероятности потери заявок в узлах СеМО могут быть рассчитаны на основе выражения (3.18) с учётом того, что  $K_1 = K_2 = 1$ :

$$\pi_1 = 1 - \frac{\rho_1}{y_1}; \quad \pi_2 = 1 - \frac{\rho_2}{y_2};$$

в этих выражениях:  $y_1 = \lambda_1 b_1$  и  $y_2 = \lambda_2 b_2$  – создаваемые в узлах нагрузки, где  $\lambda_1$  и  $\lambda_2$  – интенсивности поступления заявок в узлы 1 и 2 СеМО, для расчёта которых необходимо выполнить анализ потоков в рассматриваемой СеМО;

интенсивность  $\lambda_1$  складывается (см рис.5.16) из интенсивности  $\lambda_0$  поступления заявок из внешнего источника и интенсивности потока заявок, возвращающихся с вероятностью  $q$  в узел 1 после обслуживания в узле 2:  $\lambda_1 = \lambda_0 + q\lambda'_2$ , где  $\lambda'_2$  – рассчитанная ранее интенсивность потока выходящих из узла 2 заявок (производительность узла 2);

аналогично, из рис 5.16 можно видеть, что интенсивность  $\lambda_2$  поступающих в узел 2 заявок представляет собой интенсивность  $\lambda'_1$  потока выходящих из узла 1 заявок (производительность узла 1):  $\lambda_2 = \lambda'_1$ ;

окончательно, после некоторых преобразований выражения для расчёта вероятностей потери заявок в узлах СеМО примут вид:

$$\pi_1 = 1 - \frac{\lambda'_1}{\lambda_0 + q\lambda'_2}; \quad \pi_2 = 1 - \frac{\lambda'_2}{\lambda'_1};$$

7)

среднее время ожидания заявок в узлах рассчитывается по формулам Литтла с учётом только обслуженных заявок:

$$w_1 = l_1 / \lambda'_1; \quad w_2 = l_2 / \lambda'_2;$$

8) аналогично, среднее время пребывания заявок в узлах:

$$u_1 = m_1 / \lambda'_1 = w_1 + b; \quad u_2 = m_2 / \lambda'_2 = w_2 + b;$$

Для расчёта *сетевых характеристик* СеМО могут использоваться следующие формулы:

- 1) суммарная загрузка узлов СеМО, характеризующая среднее число одновременно работающих узлов в сети:  $\square \square \square_1 \square \square_2$ ;
- 2) суммарное число заявок в очередях:  $L \square l_1 \square l_2$ ;
- 3) суммарное число заявок в узлах:  $M \square m_1 \square m_2 \square L \square \square$ ;
- 4) производительность СеМО (интенсивность обслуженных заявок на выходе сети):  $\square \lambda'_0 = (1 - q)\lambda'_2$ ;

- 5) вероятность потери заявок в сети:  $\pi = \frac{\lambda_0 - \lambda'_0}{\lambda_0} = 1 - \frac{\lambda'_0}{\lambda_0}$ ; следует обратить

внимание, что вероятность потери заявок в сети определяется как доля потерянных заявок по отношению к поступившим в СеМО заявкам, в то время как вероятности потери  $\square_1$  и  $\square_2$  заявок в узлах СеМО определяется как доля потерянных заявок по отношению ко всем заявкам, поступившим в конкретный узел, число которых учитывает и то, что поступившая в СеМО заявка за время нахождения в сети может попасть в данный узел несколько раз.

Математические зависимости для расчёта суммарного времени ожидания заявок и времени пребывания заявок в СеМО не могут быть получены в общем виде в виду нелинейности СеМО с потерями.



## 2) Замкнутая экспоненциальная СеМО

### □ Описание замкнутой СеМО (рис.5.18).

- Сеть массового обслуживания (СеМО) – замкнутая *двухузловая*.
  - Количество приборов в узлах: узел 1 – одноканальный, узел 2 – двухканальный.
  - Поток заявок *однородный*.
  - В СеМО постоянно циркулируют  $M = 3$  заявки.

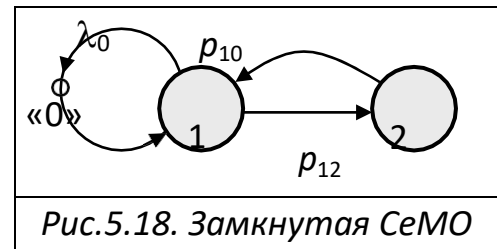


Рис.5.18. Замкнутая СеМО

### □ Предположения и допущения.

Длительности обслуживания заявок в узлах 1 и 2 распределены по экспоненциальному закону с интенсивностями  $\mu_1 = 1/b_1$  и  $\mu_2 = 1/b_2$  соответственно, где  $b_1, b_2$  средние длительности обслуживания заявок.

- Приборы в двухканальном узле 2 *идентичны* и любая заявка может обслуживаться в любом приборе.
- Заявка после обслуживания в узле 1 с вероятностью  $p_{12}$  переходит в узел 2 и с вероятностью  $p_{10} = 1 - p_{12}$  возвращается в этот же узел 1.
- Дуга, выходящая из узла 1 и входящая обратно в этот же узел, рассматривается как внешняя по отношению к СеМО, и на ней выбирается нулевая точка «0».

В замкнутой СеМО всегда существует стационарный режим, так как число заявок в сети ограничено и не может быть бесконечных очередей.

Легко убедиться, что случайный процесс, протекающий в замкнутой экспоненциальной сети, является марковским.

## Кодирование состояний марковского процесса.

Под состоянием марковского процесса будем понимать распределение заявок по узлам СеМО. Закодируем состояния следующим образом:  $(M_1, M_2)$ , где  $M_1 = \{0, 1, 2, 3\}$  – количество заявок, находящихся в узле 1 и  $M_2 = \{0, 1, 2, 3\}$  – количество заявок, находящихся в узле 2, причем суммарное число заявок в обоих узлах должно быть равно 3.

При выбранном способе кодирования система может находиться в следующих состояниях:

$E_0: (3, 0)$  – все три заявки находятся в узле 1, причем одна заявка находится на обслуживании в приборе и две заявки ожидают в накопителе;

$E_1: (2, 1)$  – две заявки находятся в узле 1 (одна на обслуживании в приборе и одна в накопителе) и одна – на обслуживании в одном из приборов узла 2;

$E_2: (1, 2)$  – одна заявка находится на обслуживании в узле 1 и две – в узле 2 (на обслуживании в обоих приборах);

$E_3: (0, 3)$  – все три заявки находятся в узле 2, причем две заявки находятся на обслуживании в обоих приборах узла 2 и одна заявка ожидает в накопителе.

### □ Размеченный граф переходов случайного процесса (рис.5.19).

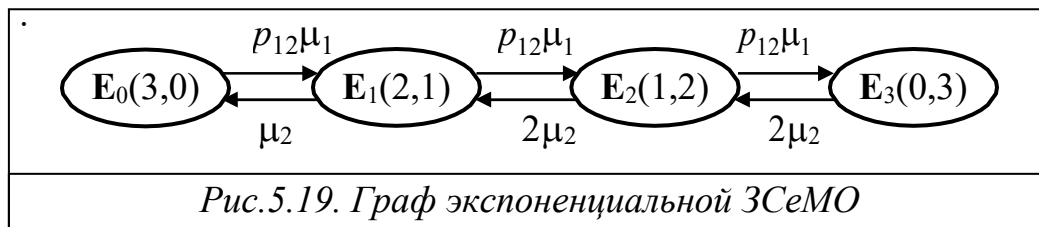
В один и тот же момент времени в замкнутой СеМО может произойти только одно из двух событий:

- 1) завершение обслуживания заявки в первом узле с интенсивностью  $\mu_1$ , при этом заявка с вероятностью  $p_{12}$  покинет этот узел и перейдет в узел 2



(интенсивность перехода  $p_{12} \mu_1$ ) или с вероятностью  $(1 - p_{12})$  останется в этом же узле, то есть состояние случайного процесса не изменится; отметим, что второй случай не отображается на графе переходов в виде дуги, выходящей из узла 1 и снова входящей в узел 1;

- 2) завершение обслуживания заявки в узле 2 с интенсивностью  $\mu_2$ , если на обслуживании в этом узле находится одна заявка (работает один прибор), или с интенсивностью  $2\mu_2$ , если на обслуживании в узле находятся две заявки (работают оба прибора); обслуженная заявка покидает этот узел и с вероятностью 1 переходит в первый узел.



■ Система уравнений.

Не составляя матрицу интенсивностей переходов, запишем систему уравнений для определения стационарных вероятностей:

$$\begin{cases} p_{12}\mu_1 p_0 = \mu_2 p_1 \\ (p_{12}\mu_1 + \mu_2) p_1 = p_{12}\mu_1 p_0 + 2\mu_2 p_2 \\ (p_{12}\mu_1 + 2\mu_2) p_2 = p_{12}\mu_1 p_1 + 2\mu_2 p_3 \\ 2\mu_2 p_3 = p_{12}\mu_1 p_2 \\ p_0 + p_1 + p_2 + p_3 = 1 \end{cases}$$

■ Расчет характеристик CeMO.

На основе полученных значений стационарных вероятностей рассчитываются узловые и сетевые характеристики CeMO с использованием следующих формул:

1. загрузка узлов:

$$\rho_1 = p_0 + p_1 + p_2; \quad \rho_2 = 0,5 p_1 + p_2 + p_3;$$

2. коэффициенты простоя узлов:

$$\eta_1 = 1 - \rho_1; \quad \eta_2 = 1 - \rho_2;$$

3. средние длины очередей заявок в узлах:

$$l_1 = 2 p_0 + p_1; \quad l_2 = p_3;$$

4. среднее число заявок в узлах:

$$m_1 = 3p_0 + 2 p_1 + p_2 \quad m_2 = p_1 + 2 p_2 + 3 p_3;$$

5. производительность замкнутой CeMO:

$$\lambda_0 = \frac{\rho_1}{\alpha_1 b_1} = \frac{\rho_2}{\alpha_2 b_2};$$

6. среднее время ожидания и пребывания заявок в узлах CeMO:

$$w_1 = \frac{l_1}{\alpha_1 \lambda_0} \quad w_2 = \frac{l_2}{\alpha_2 \lambda_0};$$

$$u_1 = \frac{l_1}{\alpha_1 \lambda_0} \quad u_2 = \frac{l_2}{\alpha_2 \lambda_0};$$

7. нагрузка в узлах сети:

$$y_1 = \alpha_1 \lambda_0 b_1; \quad y_2 = \alpha_2 \lambda_0 b_2;$$

8. среднее число параллельно работающих узлов сети, определяемое как суммарная *загрузка* всех узлов СеМО:

$$R = \rho_1 + \rho_2;$$

9. среднее число параллельно работающих *приборов* во всех узлах сети, определяемое как суммарная *нагрузка* всех узлов СеМО:

$$Y = y_1 + y_2;$$

10. суммарное число заявок во всех очередях СеМО:

$$L = l_1 + l_2;$$

11. суммарное (полное) время ожидания заявок в СеМО :

$$W = \alpha_1 w_1 + \alpha_2 w_2;$$

12. время пребывания заявок в СеМО:

$$U = \alpha_1 u_1 + \alpha_2 u_2;$$

Суммарное число заявок, циркулирующих в ЗСеМО, рассчитываемое как  $M = m_1 + m_2$ , должно совпадать с заданным числом заявок  $M = 3$ .

Следует обратить внимание на то, что временные характеристики обслуживания заявок в узлах СеМО, а, следовательно, и в сети в целом, могут быть рассчитаны только после определения производительности замкнутой СеМО, вычисляемой через найденные значения загрузок узлов.

## **Замкнутая СеМО с эрланговским обслуживанием**

### **Описание СеМО.**

- Замкнутая сеть массового обслуживания (ЗСеМО) – *двухузловая*.
- Количество приборов в узлах:  $K_1 = K_2$
- Поток заявок *однородный*
- В ЗСеМО постоянно циркулируют  $M = 3$  заявки.

Граф рассматриваемой ЗСеМО такой же, как и в предыдущем примере (рис.5.18). Отличие состоит только в том, что в рассматриваемой ЗСеМО узел 2 – одноканальный.

### **Предположения и допущения.**

- Длительность обслуживания заявок в узле 1 распределена по *закону Эрланга 2-го порядка* со средней длительностью обслуживания заявок  $b_1 = 1/\mu_1$ , а в узле 2 – по *экспоненциальному закону* со средней длительностью обслуживания заявок  $b_2 = 1/\mu_2$ , где  $\mu_1, \mu_2$  – интенсивности обслуживания заявок.

- Заявка после обслуживания в узле 1 с вероятностью  $p_{12}$  переходит в узел 2 и с вероятностью  $p_{10} = 1 - p_{12}$  возвращается в этот же узел 1.
- Дуга, выходящая из узла 1 и входящая обратно в этот же узел, рассматривается как внешняя по отношению к СеМО, и на ней отмечается нулевая точка «0».

•

### **Сведение случайного процесса к марковскому.**

Случайный процесс, протекающий в замкнутой неэкспоненциальной сети, не является марковским.

Для описания процесса функционирования такой системы в терминах марковских случайных процессов будем рассматривать функционирование системы в определенные моменты времени, в которые случайный процесс обладает марковским свойством. Для этого воспользуемся представлением случайной величины, распределенной по закону Эрланга 2-го порядка, в виде суммы двух экспоненциально распределенных случайных величин (см. раздел 2, п.2.6.1). При этом будем полагать, что обслуживание заявки в первом узле проходит две фазы, длительность каждой из которых распределена по экспоненциальному закону со средним значением  $b'_1 = b_1 / 2$ . Последнее необходимо для того, чтобы полная длительность обслуживания в узле 1 была равна  $b_1$ .

Таким образом, обслуживание заявки в СеМО можно представить как двухфазное обслуживание в первом узле и однофазное – во втором узле (рис.5.20). Длительности обслуживания в фазах  $\Phi_1$  и  $\Phi_2$  первого узла 3СеМО распределены по экспоненциальному закону с одним и тем же параметром  $\mu_1 = 1/b_1$  и с параметром  $\mu_2 = 1/b_2$  в единственной фазе второго узла. Моменты завершения обслуживания в каждой из фаз образуют цепь Маркова, так как времена нахождения в них распределены по экспоненциальному закону. Такое представление случайного процесса требует другого подхода к кодированию состояний, учитывающего распределение заявок по фазам обслуживания.

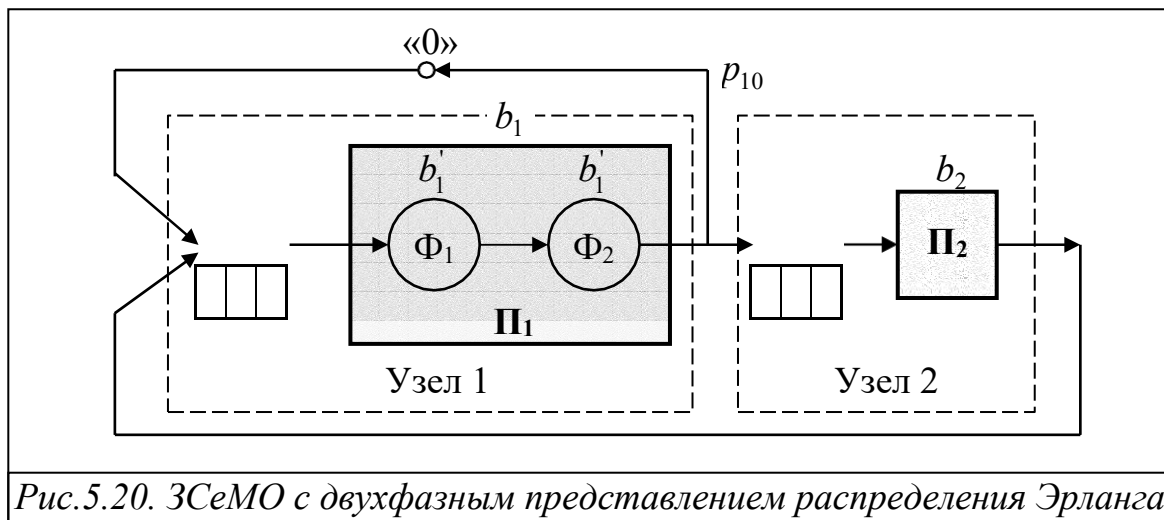


Рис.5.20. 3СеМО с двухфазным представлением распределения Эрланга

#### □ Кодирование состояний случайного процесса.

Под состоянием марковского процесса будем понимать распределение заявок по узлам СеМО с учетом того, на какой фазе обслуживания распределения Эрланга находится заявка в узле 1.

Для этого закодируем состояния следующим образом:  $(M_1, M_2)$ , где  $M_1 = \{0, 1_1, 1_2, 2_1, 2_2, 3\}$  – количество заявок, находящихся в узле 1 (индексы отражают нахождение заявки на 1-й или 2-й фазе распределения Эрланга), и  $M_2 = \{0, 1, 2, 3\}$  – количество заявок, находящихся в узле 2, причем суммарное число заявок в обоих узлах должно быть равно 3.

При выбранном способе кодирования система может находиться в следующих состояниях:

$E_1: (3_1, 0)$  – все три заявки находятся в узле 1, причем одна заявка находится на обслуживании в приборе на *первой фазе*, и две заявки ожидают в накопителе;

$E_2: (3_2, 0)$  – все три заявки находятся в узле 1, причем одна заявка находится

на обслуживании в приборе на *второй фазе*, и две заявки ожидают в накопителе;

$E_3: (2_1, 1)$  – две заявки находятся в узле 1 (одна на обслуживании в приборе на *первой фазе* и одна в накопителе) и одна – на обслуживании в узле 2;

$E_4: (2_2, 1)$  – две заявки находятся в узле 1 (одна на обслуживании в приборе на *второй фазе* и одна в накопителе) и одна – на обслуживании в узле 2;

$E_5: (1_1, 2)$  – одна заявка находится в узле 1 на обслуживании в приборе на *первой фазе* и две заявки находятся в узле 2, причем одна из них находится на обслуживании в приборе, а вторая заявка ожидает в накопителе;

$E_6: (1_2, 2)$  – одна заявка находится в узле 1 на обслуживании в приборе на *второй фазе* и две заявки находятся в узле 2, причем одна из них находится на обслуживании в приборе, а вторая заявка ожидает в накопителе;

$E_7: (0, 3)$  – все три заявки находятся в узле 2, причем одна заявка находится на обслуживании в приборе, а две другие – ожидают в накопителе.

■ Размеченный граф переходов случайного процесса.

На рис.5.21 представлен граф переходов марковского процесса для рассматриваемой неэкспоненциальной СеМО. Для понимания процесса составления графа переходов вместо номеров состояний в вершинах графа указаны коды состояний.

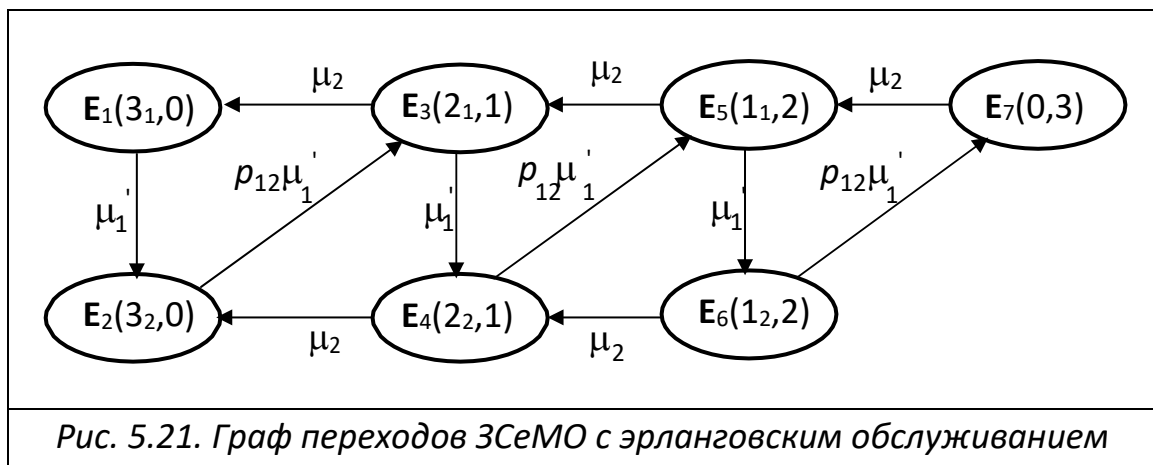


Рис. 5.21. Граф переходов 3СеМО с эрланговским обслуживанием

Из состояния  $E_1=(3_1, 0)$  переход возможен только в одно состояние  $E_2=(3_2, 0)$  с интенсивностью  $\mu_1$  обслуживания на первой фазе, поскольку все заявки в первом узле обязательно проходят две фазы обслуживания.

Из состояния  $E_2=(3_2, 0)$  переход возможен также только в одно состояние  $E_3=(2_1, 1)$ . Это соответствует завершению обслуживания на второй фазе заявки в узле 1 (с интенсивностью  $\mu_1'$ ) и ее передаче в узел 2 (с вероятностью  $p_{12}$ ). Отсюда интенсивность перехода марковского процесса в состояние  $E_3=(2_1, 1)$  будет равна произведению  $p_{12}\mu_1'$ . Заметим, что с вероятностью  $1-p_{12}$  марковский процесс останется в том же состоянии, что соответствует возврату заявки в узел 1.

Из состояния  $E_3=(2_1, 1)$  переход возможен в одно из двух состояний:

- в состояние  $E_4=(2_2, 1)$ , что соответствует завершению обслуживания заявки на первой фазе в узле 1 (с интенсивностью  $\mu_1'$ ) и переходу к обслуживанию на второй фазе в том же узле 1;
- в состояние  $E_1=(3_1, 0)$ , что соответствует завершению обслуживания заявки в узле 2 (с интенсивностью  $\mu_2$ ) и ее передаче в узел 1 (с вероятностью 1).
- Следует помнить, что в любой момент времени может произойти только одно событие. Вероятность двух и более событий пренебрежимо мала.

Поэтому из состояния  $E_3=(21, 1)$  не возможен переход в состояние  $E_3=(32, 0)$ , означающий завершение обслуживания заявки на первой фазе в узле 1 и одновременное завершение обслуживания заявки в узле 2.

Аналогичные рассуждения справедливы для состояний  $E_4=(2_2, 1)$ ,  $E_5=(1_1, 2)$  и  $E_6=(1_2, 2)$ .

Из состояния  $E_7=(0, 3)$  переход возможен только в состояние  $E_5=(1_1, 2)$ , означающий завершение обслуживания заявки в узле 2 и ее передачу на обслуживание в узел 1, причем обслуживание новой заявки всегда начинается на первой фазе. По этой причине переход в состояние  $E_6=(1_2, 2)$  не возможен.

### Система уравнений.

Не составляя матрицу интенсивностей переходов, запишем систему уравнений для определения стационарных вероятностей состояний:

$$\begin{cases} \mu_1' p_1 = \mu_2 p_3; \\ p_{12} \mu_1' p_2 = \mu_1' p_1 + \mu_2 p_4; \\ (\mu_1' + \mu_2) p_3 = p_{12} \mu_1' p_2 + \mu_2 p_5; \\ (p_{12} \mu_1' + \mu_2) p_4 = \mu_1' p_3 + \mu_2 p_6; \\ (\mu_1' + \mu_2) p_5 = p_{12} \mu_1' p_4 + \mu_2 p_7; \\ (p_{12} \mu_1' + \mu_2) p_6 = \mu_1' p_5; \\ \mu_2 p_7 = p_{12} \mu_1' p_6; \\ p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1; \end{cases}$$

□ Расчет характеристик СеМО.

Характеристики ЗСеМО определяются в такой последовательности:

1) загрузка и коэффициенты простоя узлов:

$$\rho_1 = p_1 + p_2 + p_3 + p_4 + p_5 + p_6; \quad \rho_2 = p_3 + p_4 + p_5 + p_6 + p_7;$$

$$\eta_1 = 1 - \rho_1; \quad \eta_2 = 1 - \rho_2;$$

2) среднее число параллельно работающих узлов сети, определяемое как суммарная загрузка всех узлов СеМО:

$$R = \rho_1 + \rho_2;$$

3) среднее число заявок в очередях и в узлах СеМО:

$$l_1 = 2(p_1 + p_2) + p_3 + p_4; \quad l_2 = p_5 + p_6 + 2 p_7;$$

$$m_1 = 3(p_1 + p_2) + 2(p_3 + p_4) + p_5 + p_6; \quad m_2 = p_3 + p_4 + 2(p_5 + p_6) + 3 p_7;$$

4) суммарное число заявок во всех очередях СеМО:

$$L = l_1 + l_2;$$

5) производительность замкнутой СеМО:

$$\lambda_0 = \frac{\rho_1}{\alpha_1 b_1} = \frac{\rho_2}{\alpha_2 b_2};$$

где  $\alpha_1$  и  $\alpha_2$  - коэффициенты передач соответственно узла 1 и узла 2;

6) средние времена ожидания и пребывания заявок в узлах СеМО

$$w_1 = \frac{l_1}{\alpha_1 \lambda_0} \quad w_2 = \frac{l_2}{\alpha_2 \lambda_0};$$

$$u_1 = \frac{l_1}{\alpha_1 \lambda_0} \quad u_2 = \frac{l_2}{\alpha_2 \lambda_0};$$

7) суммарное (полное) время ожидания и время пребывания заявок в

СеМО:

$$W = \alpha_1 w_1 + \alpha_2 w_2;$$

$$U = \alpha_1 u_1 + \alpha_2 u_2;$$

8) нагрузка в узлах сети:

$$y_1 = \alpha_1 \lambda_0 b_1; \quad y_2 = \alpha_2 \lambda_0 b_2;$$

9) среднее число параллельно работающих *приборов* во всех узлах сети, определяемое как суммарная *нагрузка* всех узлов СеМО:

$$Y = y_1 + y_2;$$

Суммарное число заявок, циркулирующих в СеМО, рассчитываемое как  $M = m_1 + m_2$ , должно совпадать с заданным числом заявок в замкнутой сети:  $M = 3$ .

### Замкнутая СеМО с гиперэкспоненциальным обслуживанием

#### □ Описание СеМО.

- Сеть массового обслуживания (СеМО) – *двухузловая*.
- Количество приборов в узлах:  $K_1 = K_2 = 1$ .
- Поток заявок *однородный*.
- В СеМО постоянно циркулируют  $M=3$  заявки.

#### □ Предположения и допущения.

- Длительность обслуживания заявок в узле 1 распределена по *гиперэкспоненциальному закону* со средней длительностью обслуживания заявок  $b_1 = 1/\mu_1$  и коэффициентом вариации  $v=2$ , а в узле 2 – по экспоненциальному закону со средней длительностью обслуживания заявок  $b_2 = 1/\mu_2$ , где  $\mu_1, \mu_2$  – интенсивности обслуживания заявок
  - Заявка после обслуживания в узле 1 с вероятностью  $p_{12}$  переходит в узел 2 и с вероятностью  $p_{10}=1-p_{12}$  возвращается в этот же узел 1.
  - Дуга, выходящая из узла 1 и входящая обратно в этот же узел, рассматривается как внешняя по отношению к СеМО, и на ней выбирается нулевая точка «0».

В замкнутой СеМО всегда существует стационарный режим.

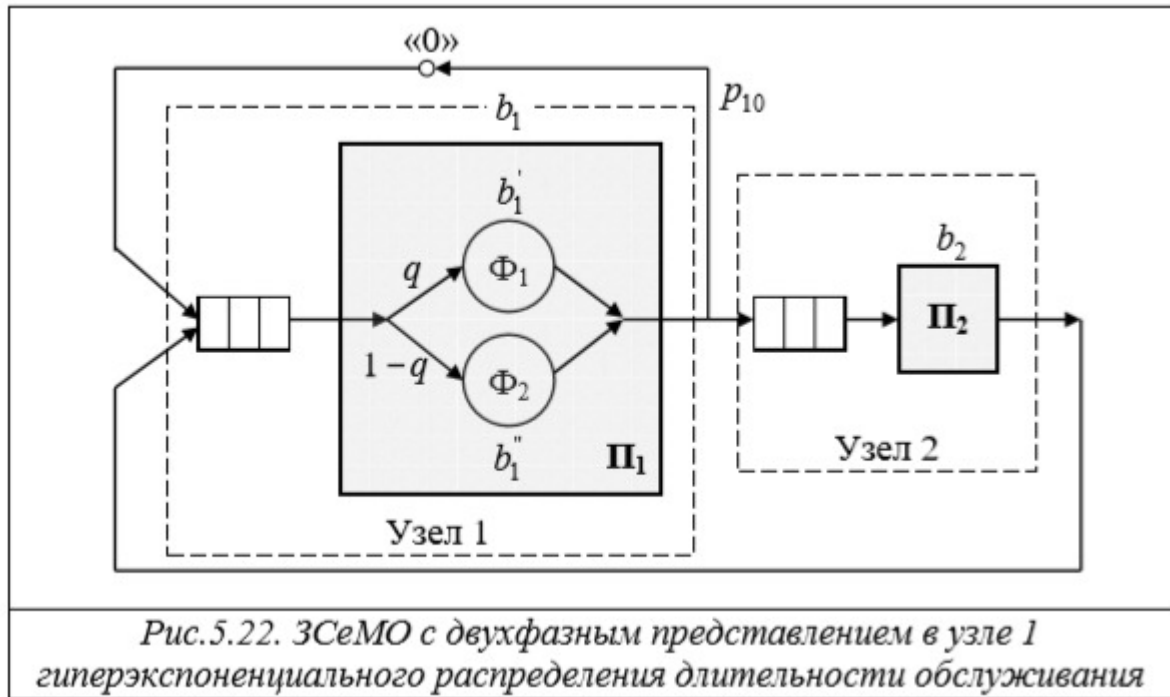
### Сведение случайного процесса к марковскому.

Для описания процесса функционирования в замкнутой неэкспоненциальной сети в терминах марковских случайных процессов, как и ранее будем рассматривать функционирование системы в определенные моменты времени, в которые случайный процесс обладает марковским свойством. Для этого воспользуемся представлением случайной величины, распределенной по гиперэкспоненциальному закону, в виде композиции двух экспоненциально распределенных случайных величин (см. раздел 2, п.2.6.2), каждая из которых появляется с вероятностями  $q$  и  $(1-q)$  соответственно. В первом узле 3СеМО такое представление реализуется в виде двух параллельных экспоненциальных фаз, обслуживающих заявки по следующей схеме (рис 5.22).

- заявка с вероятностью  $q=0,1$  попадает на обслуживание в первую фазу, длительность обслуживания в которой распределена по экспоненциальному закону со средним значением  $b'$ , после чего покидает узел.



- заявка с вероятностью  $(1 - q) = 0,9$  попадает на обслуживание во вторую фазу, длительность обслуживания в которой распределена по экспоненциальному закону со средним значением  $b_1''$  после чего покидает первый узел.
- значения длительностей обслуживания в этих двух фазах таковы, что выполняется условие  $qb_1 + (1-q)b_1'' = b_1$ . Последнее необходимо для того, чтобы средняя длительность обслуживания в узле 1 была равна  $b_1$ .



Моменты завершения обслуживания в каждой из фаз образуют цепь Маркова, так как времена нахождения в них распределены по экспоненциальному закону.

### Кодирование состояний случайного процесса.

Под состоянием марковского процесса будем понимать распределение заявок по узлам СеМО с учетом того, на какой фазе обслуживания в узле 1 находится заявка.

Для этого закодируем состояния следующим образом:  $(\mathbf{M}_1, \mathbf{M}_2)$ , где  $\mathbf{M}_1 = \{0, 1_1, 1_2, 2_1, 2_2, 3\}$  – количество заявок, находящихся в узле 1 (индексы отражают нахождение заявки на 1-й или 2-й фазе гиперэкспоненциального распределения), и  $\mathbf{M}_2 = \{0, 1, 2, 3\}$  – количество заявок, находящихся в узле 2, причем суммарное число заявок в обоих узлах должно быть равно 3.

При выбранном способе кодирования система может находиться, как и в предыдущем примере, в следующих состояниях:

$E_1: (3_1, 0)$  – все три заявки находятся в узле 1, причем одна заявка находится на обслуживании в приборе на *первой фазе*, и две заявки ожидают в накопителе;

$E_2: (3_2, 0)$  – все три заявки находятся в узле 1, причем одна заявка находится на обслуживании в приборе на *второй фазе*, и две заявки ожидают в накопителе;

$E_3: (2_1, 1)$  – две заявки находятся в узле 1 (одна на обслуживании в приборе на *первой фазе* и одна в накопителе) и одна – на обслуживании в узле 2;

$E_4: (2_2, 1)$  – две заявки находятся в узле 1 (одна на обслуживании в приборе на *второй фазе* и одна в накопителе) и одна – на обслуживании в узле 2;

$E_5: (1_1, 2)$  – одна заявка находится в узле 1 на обслуживании в приборе на *первой фазе* и две заявки находятся в узле 2, причем одна из них находится на обслуживании в приборе, а вторая заявка ожидает в накопителе;

$E_6: (1_2, 2)$  – одна заявка находится в узле 1 на обслуживании в приборе на *второй фазе* и две заявки находятся в узле 2, причем одна из них находится на обслуживании в приборе, а вторая заявка ожидает в накопителе;

$E_7: (0, 3)$  – три заявки находятся в узле 2, причем одна заявка – на обслуживании в приборе, а две другие – ожидают в накопителе.

### Размеченный граф переходов случайного процесса.

На рис.5.23 представлен граф переходов марковского процесса для рассматриваемой неэкспоненциальной СеМО с гиперэкспоненциальным распределением длительности обслуживания заявок в первом узле. Для понимания процесса составления графа переходов вместо номеров состояний в вершинах графа указаны коды состояний, а для того чтобы не загромождать рисунок, используются следующие обозначения для интенсивностей переходов:  $g_1 = (1-q)(1-p_{12})\mu'_1$ ;  $g_2 = (1-q)(1-p_{12})\mu'_1$ ;  $g_3 = (1-q)(1-p_{12})\mu''_1$ ;  $g_4 = qp_{12}\mu''_1$ .

Рассмотрим подробно все возможные переходы для каждого состояния  $E_i$  ( $i = 1, \dots, 7$ ) марковского случайного процесса.

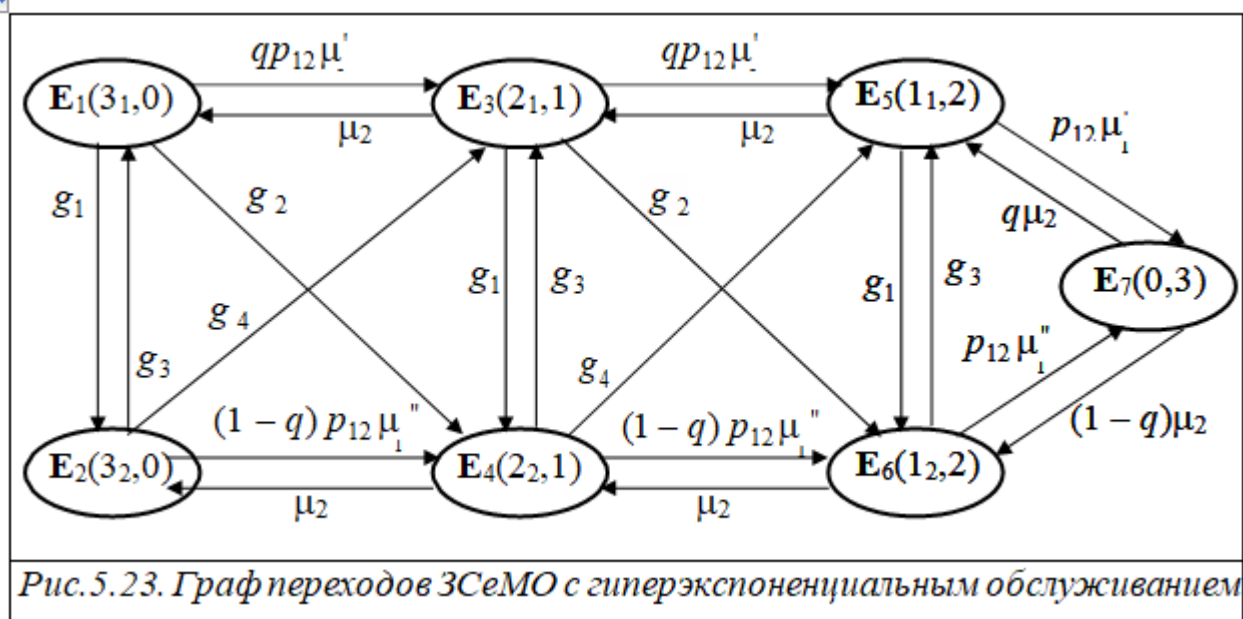


Рис.5.23. Граф переходов 3СеМО с гиперэкспоненциальным обслуживанием

(а) Состояние  $E_1$ . Если случайный процесс находится в состоянии

$E_1=(3_1, 0)$ , то по завершению обслуживания заявки случайный процесс может перейти в одно из трёх состояний:  $E_2=(3_2, 0)$ ,  $E_3=(2_1, 1)$  и  $E_4=(2_2, 1)$  или остаться в том же состоянии. Напомним, что если случайный процесс остаётся в том же состоянии, то это никак не отображается на графе переходов.

Случайный процесс перейдёт из состояния  $E_1=(3_1, 0)$  в состояние  $E_2=(3_2, 1)$  при выполнении следующих условий:

- завершится обслуживание заявки, находящейся на обслуживании в фазе  $\Phi_1$ ; интенсивность этого события  $\mu'_1 = 1/b'_1$ ;
- заявка, завершившая обслуживание в узле 1, вернётся в этот же узел



и встанет в конец очереди; вероятность этого события равна  $p_{10} = 1 - p_{12}$ ;

- в узле 1 очередная заявка, которая поступит на обслуживание из очереди в прибор  $\Pi_1$ , попадёт на обслуживание в фазу  $\Phi_2$ ; вероятность этого события равна  $(1 - q)$ .

Таким образом, интенсивность перехода из состояния  $E_1=(3_1, 0)$  в состояние  $E_2=(3_2, 0)$  будет равна  $g_1 = (1 - q)(1 - p_{12})\mu_1'$

Случайный процесс перейдёт из состояния  $E_1=(3_1, 0)$  в состояние  $E_3=(2_1, 1)$  при выполнении следующих условий:

- завершится обслуживание заявки, находящейся на обслуживании в фазе  $\Phi_1$ ; интенсивность этого события  $\mu_1' = 1/b_1'$ ;

- заявка, завершившая обслуживание в узле 1, перейдёт в узел 2; вероятность этого события равна  $p_{12}$ ;

- в узле 1 новая заявка, которая поступит на обслуживание из очереди в прибор  $\Pi_1$ , попадёт на обслуживание в фазу  $\Phi_1$ ; вероятность этого события –  $q$ .

Таким образом, интенсивность перехода из состояния  $E_1=(3_1, 0)$  в состояние  $E_3=(2_1, 1)$  будет равна  $qp_{12}\mu_1'$ .

Случайный процесс перейдёт из состояния  $E_1=(3_1, 0)$  в состояние  $E_4=(2_2, 1)$  при выполнении следующих условий:

- завершится обслуживание заявки, находящейся на обслуживании в фазе  $\Phi_1$ ; интенсивность этого события  $\mu_1' = 1/b_1'$ ;

- заявка, завершившая обслуживание в узле 1, перейдёт в узел 2; вероятность этого события равна  $p_{12}$ ;

- в узле 1 новая заявка, которая поступит на обслуживание из очереди в прибор  $\Pi_1$ , попадёт на обслуживание в фазу  $\Phi_2$ ; вероятность этого события –  $(1 - q)$ .

Таким образом, интенсивность перехода из состояния  $E_1=(3_1, 0)$  в состояние  $E_4=(2_2, 1)$  будет равна  $g_2 = (1 - q)p_{12}\mu_1'$ .

**Состояние  $E_2$**  Случайный процесс из состояния  $E_2=(3_2, 0)$  по завершению обслуживания заявки также может перейти в одно из трёх состояний:  $E_1=(3_1, 0)$ ,  $E_3=(2_1, 1)$  и  $E_4=(2_2, 1)$  или остаться в том же состоянии.

Случайный процесс перейдёт из состояния  $E_2=(3_2, 0)$  в состояние  $E_1=(3_1, 1)$  при выполнении следующих условий:

- с интенсивностью  $\mu_1'' = 1/b_1''$  завершится обслуживание заявки в фазе  $\Phi_2$ ;

- с вероятностью  $p_{10} = 1 - p_{12}$  заявка, завершившая обслуживание в узле 1, вернётся в этот же узел и встанет в конец очереди;

- с вероятностью  $q$  в узле 1 очередная заявка, которая поступит из очереди в прибор  $\Pi_1$ , попадёт на обслуживание в фазу  $\Phi_1$ .

Таким образом, интенсивность перехода из состояния  $E_1=(3_1, 0)$  в состояние  $E_2=(3_2, 0)$  будет равна  $g_3 = q(1 - p_{12})\mu_1''$

Случайный процесс перейдёт из состояния  $E_2=(3_2, 0)$  в состояние  $E_3=(2_1, 1)$  при выполнении следующих условий:

- с интенсивностью  $\mu_1'' = 1/b_1''$  завершится обслуживание заявки в фазе

$\Phi_2$ ;

- с вероятностью  $p_{12}$  заявка, завершившая обслуживание в узле 1, перейдет в узел 2;
- с вероятностью  $q$  в узле 1 очередная заявка, которая поступит из очереди в прибор  $\Pi_1$ , попадет на обслуживание в фазу  $\Phi_1$ .

Таким образом, интенсивность перехода  $E_2=(3_2, 0)$  в  $E_3=(2_1, 1)$  будет равна  $g_4 = q p_{12} \mu_1$

Случайный процесс перейдет из состояния  $E_2=(3_2, 0)$  в состояние  $E_4=(2_2, 1)$  при выполнении следующих условий:

- с интенсивностью  $u_1'' = 1/b_1''$  завершится обслуживание заявки в фазе  $\Phi_2$ ;
- с вероятностью  $p_{12}$  заявка, завершившая обслуживание в узле 1, перейдет в узел 2;
- с вероятностью  $(1-q)$  в узле 1 очередная заявка, которая поступит из очереди в прибор  $\Pi_1$ , попадет на обслуживание в фазу  $\Phi_2$ .

Таким образом, интенсивность перехода из  $E_2=(3_2, 0)$  в  $E_4=(2_2, 1)$  будет равна  $(1-q) p_{12} \mu_1$ .

**Состояния  $E_3$  и  $E_4$ .** Если случайный процесс находится в состоянии  $E_3=(2_1, 1)$  или  $E_4=(2_2, 1)$ , то кроме аналогичных переходов, связанных с завершением обслуживания заявки в узле 1, имеется ещё один переход в состояния  $E_1=(3_1, 0)$  и  $E_2=(3_2, 0)$  соответственно, связанный с завершением обслуживания заявки в узле 2. Интенсивность перехода из  $E_3=(2_1, 1)$  в  $E_1=(3_1, 0)$  и из  $E_4=(2_2, 1)$  в  $E_2=(3_2, 0)$  равна интенсивности обслуживания  $\mu_2$  в узле 2. Отметим, что переходы из  $E_3=(2_1, 1)$  в  $E_2=(3_2, 0)$  и из  $E_4=(2_2, 1)$  в  $E_1=(3_1, 0)$  отсутствуют, так как заявка, находящаяся на обслуживании в первом узле, остаётся в той же фазе обслуживания, которая была в момент завершения обслуживания заявки в узле 2. Это является следствием того, что в случайных процессах с непрерывным временем вероятность одновременного появления двух событий (завершение обслуживания в узле 1 и в узле 2) равна нулю.

**Состояния  $E_5$  и  $E_6$ .** Переходы из состояний  $E_5=(1_1, 2)$  и  $E_6=(1_2, 2)$  аналогичны переходам из  $E_3=(2_1, 1)$  и  $E_4=(2_2, 1)$  за исключением переходов в состояние  $E_7=(0, 3)$ . Интенсивности переходов из  $E_5=(1_1, 2)$  и  $E_6=(1_2, 2)$  в  $E_7=(0, 3)$  определяются как произведение интенсивности обслуживания в соответствующей фазе узла 1 на вероятность того, что заявка, завершившая обслуживание в узле 1, перейдет в узел 2:  $p_{12} \mu_1$  и  $p_{12} \mu_1$ .

**Состояние  $E_7$ .** Переходы из состояния  $E_7=(0, 3)$  связаны с завершением обслуживания с интенсивностью  $\mu_2$  заявки в узле 2, которая переходит в узел 1 и с вероятностью  $q$  попадает на обслуживание в фазу  $\Phi_1$  или с вероятностью  $(1-q)$  – в фазу  $\Phi_2$ . Соответственно интенсивности переходов будут равны  $q \mu_2$  и  $(1-q) \mu_2$

### Расчет характеристик СеМО.

Не составляя матрицу интенсивностей переходов и не выписывая систему линейных алгебраических уравнений для определения стационарных

вероятностей состояний, приведём математические зависимости для расчёта характеристик функционирования ЗСеМО:

1) загрузка и коэффициенты простоя узлов:

$$\rho_1 = p_1 + p_2 + p_3 + p_4 + p_5 + p_6; \rho_2 = p_3 + p_4 + p_5 + p_6 + p_7;$$

$$\eta_1 = 1 - \rho_1; \eta_2 = 1 - \rho_2;$$

2) среднее число параллельно работающих узлов сети, определяемое как суммарная загрузка всех узлов СеМО:

$$R = \rho_1 + \rho_2;$$

3) среднее число заявок в очередях и в узлах СеМО:

$$l_1 = 2(p_1 + p_2) + p_3 + p_4; \quad l_2 = p_5 + p_6 + 2p_7;$$

$$m_1 = 3(p_1 + p_2) + 2(p_3 + p_4) + p_5 + p_6;$$

$$m_2 = p_3 + p_4 + 2(p_5 + p_6) + 3p_7;$$

4) суммарное число заявок во всех очередях СеМО:

$$L = l_1 + l_2;$$

5) производительность замкнутой СеМО:

$$\lambda_0 = \frac{\rho_1}{\alpha_1 b_1} = \frac{\rho_2}{\alpha_2 b_2};$$

где  $\alpha_1$  и  $\alpha_2$  – коэффициенты передачи соответственно узла 1 и узла 2;

6) средние времена ожидания и пребывания заявок в узлах СеМО:

$$w_1 = \frac{l_1}{\alpha_1 \lambda_0}; \quad w_2 = \frac{l_2}{\alpha_2 \lambda_0};$$

$$u_1 = \frac{l_1}{\alpha_1 \lambda_0}; \quad u_2 = \frac{l_2}{\alpha_2 \lambda_0};$$

7) суммарное (полное) время ожидания и время пребывания заявок в узлах СеМО:

$$W = \alpha_1 w_1 + \alpha_2 w_2;$$

$$U = \alpha_1 u_1 + \alpha_2 u_2;$$

8) нагрузка в узлах сети:

$$y_1 = \alpha_1 \lambda_0 b_1; \quad y_2 = \alpha_2 \lambda_0 b_2;$$

9) среднее число параллельно работающих приборов во всех узлах сети, определяемое как суммарная нагрузка всех узлов СеМО:

$$Y = y_1 + y_2.$$

Суммарное число заявок, циркулирующих в СеМО, рассчитываемое как должно совпадать с заданным числом заявок в замкнутой сети  $M = 3$

**Задание на самостоятельную работу:**

1. По графу переходов рис.5.23 построить матрицу интенсивностей переходов и составить систему линейных алгебраических уравнений для расчёта стационарных вероятностей состояний.

2. Выполнить детальный анализ свойств исследуемой системы.



## ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СМО и СеМО

### Основы имитационного моделирования

**Статистическое моделирование** – метод исследования сложных систем, основанный на описании процессов функционирования отдельных элементов в их взаимосвязи с целью получения множества частных результатов, подлежащих обработке методами математической статистики для получения конечных результатов. В основе статистического моделирования лежит метод статистических испытаний – метод Монте-Карло.

**Имитационная модель** – универсальное средство исследования сложных систем, представляющее собой логико-алгоритмическое описание поведения отдельных элементов системы и правил их взаимодействия, отображающих последовательность событий, возникающих в моделируемой системе.

Если статистическое моделирование выполняется с использованием имитационной модели, то такое моделирование называется *имитационным*.

Понятия «статистическое и имитационное моделирование» часто рассматривают как синонимы. Однако следует иметь в виду, что статистическое моделирование не обязательно является имитационным. Например, вычисление определённого интеграла методом Монте-Карло путем определения подынтегральной площади на основе множества статистических испытаний, относится к статистическому моделированию, но не может называться имитационным.

Наиболее широкое применение имитационное моделирование получило при исследовании сложных систем с дискретным характером функционирования, в том числе моделей массового обслуживания. Для описания процессов функционирования таких систем обычно используются временные диаграммы.

**Временная диаграмма** – графическое представление последовательности событий, происходящих в системе. Для построения временных диаграмм необходимо достаточно четко представлять взаимосвязь событий внутри системы. Степень детализации при составлении диаграмм зависит от свойств моделируемой системы и от целей моделирования.

Поскольку функционирование любой системы достаточно полно отображается в виде временной диаграммы, *имитационное моделирование можно рассматривать как процесс реализации диаграммы функционирования исследуемой системы на основе сведений о характере функционирования отдельных элементов и их взаимосвязи.*

Имитационное моделирование обычно проводится на ЭВМ в соответствии с программой, реализующей заданное конкретное логико- алгоритмическое описание. При этом несколько часов, недель или лет работы исследуемой системы могут быть промоделированы на ЭВМ за несколько минут. В большинстве случаев модель является не точным аналогом системы, а скорее её символическим отображением. Однако такая модель позволяет производить измерения, которые невозможно произвести каким-либо другим способом.

Имитационное моделирование обеспечивает возможность испытания, оценки и проведения экспериментов с исследуемой системой без каких-либо непосредственных воздействий на нее.

Первым шагом при анализе любой конкретной системы является выделение элементов, и формулирование логических правил, управляющих взаимодействием этих элементов. Полученное в результате этого описание называется **моделью системы**. Модель обычно включает в себя те аспекты системы, которые представляют интерес или нуждаются в исследовании.

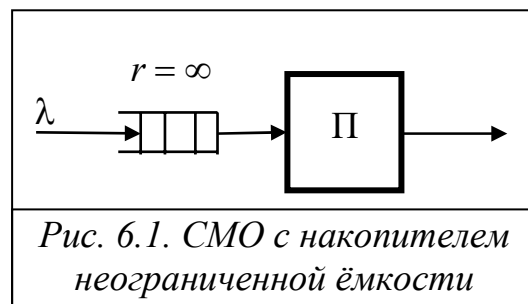
Поскольку целью построения любой модели является исследование характеристик моделируемой системы, в имитационную модель должны быть включены средства сбора и обработки статистической информации по всем интересующим характеристикам, основанные на методах математической статистики.

### Принципы организации имитационного моделирования

Рассмотрим принципы имитационного моделирования на примере простейшей базовой модели в виде одноканальной системы массового обслуживания с однородным потоком заявок (рис.6.1), в которую поступает случайный поток заявок с интервалами между соседними заявками, распределёнными по закону  $A(\tau)$ , а длительность обслуживания

заявок в приборе распределена по закону  $B(\tau)$ .

Процесс функционирования такой системы может быть представлен в виде временных диаграмм, на основе которых могут быть измерены и рассчитаны характеристики обслуживания заявок. Поскольку процессы поступления и обслуживания заявок в системе носят случайный характер, то для построения диаграмм необходимо иметь *генераторы случайных чисел*.



Положим, что в нашем распоряжении имеются генераторы случай-

ных чисел, формирующие значения соответствующих случайных величин с заданными законами распределений  $A(\tau)$  и  $B(\tau)$ . Тогда можно построить временные диаграммы, отображающие процесс функционирования рассматриваемой системы.

На рис. 6.2 представлены четыре диаграммы, отображающие

- **«процесс поступления заявок»** в виде моментов  $t_i$  поступления заявок в систему, формируемых по правилу:  $t_i = t_{i-1} + \tau_{ai}$ , ( $t_0=0$ ), где  $\tau_{ai}$  ( $i=1,2,..$ ) – интервалы между поступающими в систему заявками, значения которых вырабатываются с помощью генератора случайных величин  $A(\tau)$ .

- **«процесс обслуживания в приборе»**, представленный в виде длительностей обслуживания  $\tau_{bi}$ , которых вырабатываются с помощью генератора случайных величин  $B(\tau)$ , и моментов завершения обслуживания  $t'_i$  заявок в приборе, определяемых по следующему правилу:

$t'_i = t_i + \tau_{bi}$ , если на момент поступления  $i$ -й заявки обслуживающий прибор был свободен;

$t'_i = t'_{i-1} + \tau_{bi}$ , если на момент поступления  $i$ -й заявки обслуживающий прибор был занят обслуживанием предыдущей заявки ( $i = 1, 2, \dots; t'_0 = 0$ );

- **«модельное или реальное время»**, показывающее дискретное (скачкообразное) изменение времени в реальной системе, каждый момент которого соответствует одному из следующих событий: поступление заявки в систему или завершение обслуживания заявки в приборе; отметим, что в эти моменты времени происходит изменение состояния системы, описываемое числом заявок, находящихся в системе;

- **«число заявок в системе»**, описывающее состояние дискретной системы и изменяющееся по правилу: увеличение на 1 в момент поступления заявки в систему и уменьшение на 1 в момент завершения обслуживания.

При соблюдении выбранного временного масштаба представленные диаграммы позволяют путем измерения определить значения вероятностно-временных характеристик функционирования моделируемой системы, в частности, как показано на второй диаграмме, время нахождения (пребывания) каждой заявки в системе  $\tau_{ui}$  ( $i = 1, 2, \dots$ ).

Очевидно, что время пребывания заявок в системе – величина случайная. В простейшем случае, применяя методы математической статистики, можно рассчитать два первых момента распределения времени пребывания:

- математическое ожидание:

$$u = \frac{1}{N} \sum_{i=1}^N \tau_{ui};$$

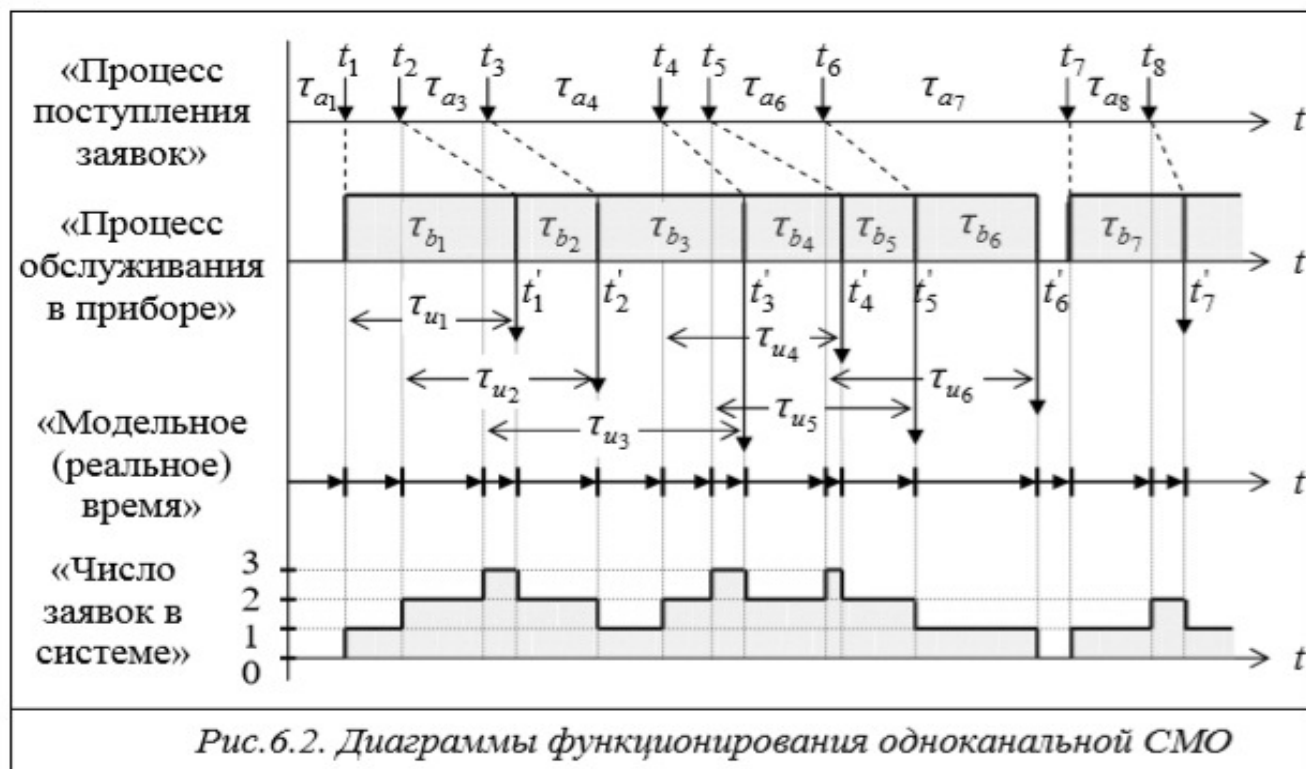
- второй начальный момент:

$$u^{(2)} = \frac{1}{N-1} \sum_{i=1}^N \tau_{ui}^2;$$

где  $N$  - количество значений времени пребывания заявок, полученных на диаграмме, то есть количество заявок, отображенных на диаграмме как прошедшие через систему и покинувшие её.

Отсюда легко могут быть получены значения дисперсии,

среднеквадратического отклонения и коэффициента вариации времени пребывания заявок в системе.



На основе полученных с помощью временных диаграмм значений времени пребывания заявок в системе можно построить гистограмму функции или плотности распределения времени пребывания.

Точность полученных числовых моментов распределения и качество гистограмм существенно зависит от количества значений  $N$  времени пребывания заявок, на основе которых они рассчитываются: чем больше  $N$ , тем точнее результаты расчета. Значение  $N$  может составлять от нескольких тысяч до десятков миллионов. Конкретное значение  $N$  зависит от многих факторов, влияющих на скорость сходимости результатов к истинному значению, основными среди которых при моделировании систем и сетей массового обслуживания являются законы распределений интервалов между поступающими заявками и длительностей обслуживания, загрузка системы, сложность модели, количество классов заявок и т.д.

Ясно, что построение вручную таких временных диаграмм с тысячами и более проходящими через систему заявками, нереально. В то же время, использование ЭВМ для реализации временных диаграмм позволяет существенно ускорить процессы моделирования и получения конечного результата. Поэтому, как сказано выше, имитационное моделирование можно рассматривать как процесс реализации диаграммы функционирования исследуемой системы.

Таким образом, имитационная модель представляет собой алгоритм реализации временной диаграммы функционирования исследуемой системы. Наличие встроенных в большинство алгоритмических языков генераторов случайных чисел значительно упрощает процесс реализации имитационной модели на ЭВМ. Однако при этом остаётся ряд проблем, требующих своего решения. Одна из них заключается в принципе реализации временной диаграммы и, связанной с ней, проблемой организации службы времени в имитационной модели.

В простейшем случае временная диаграмма может быть реализована



следующим образом: сначала формируются моменты поступления всех заявок в систему, а затем для каждой заявки определяются длительности обслуживания в приборе и формируются моменты завершения обслуживания (выхода заявок из системы). Очевидно, что такой подход неприемлем, поскольку даже для нашей очень простой системы придётся хранить в памяти ЭВМ одновременно миллионы значений моментов поступления и завершения обслуживания заявок, а также других переменных, причём с увеличением количества классов заявок и количества обслуживающих приборов это число увеличится многократно.

Второй подход, который может быть предложен для реализации временной диаграммы, – пошаговое построение диаграммы. Для этого следует сформировать переменную для модельного времени и выбрать шаг  $\Delta t$  его изменения. В каждый такой момент времени необходимо проверять, какое событие (поступление в систему или завершение обслуживания заявки) произошло в системе за предыдущий интервал  $\Delta t$ .

Этот подход значительно сокращает потребность в памяти, поскольку в этом случае в каждый момент времени необходимо хранить в памяти ЭВМ значения параметров (моментов поступления и завершения обслуживания) только тех заявок, которые находятся в системе на данный момент времени.

Недостатки такого подхода очевидны. Во-первых, проблематичным является выбор длины интервала  $\Delta t$ . С одной стороны, интервал  $\Delta t$  должен быть как можно меньше для уменьшения методической погрешности моделирования, с другой стороны, интервал  $\Delta t$  должен быть как можно больше для уменьшения времени моделирования.

Наиболее эффективным подходом признан подход с *переменным шагом продвижения модельного времени*, который реализуется в соответствии с принципом «до ближайшего события». **Принцип «продвижения модельного времени до ближайшего события»** заключается в следующем. По всем процессам, параллельно протекающим в исследуемой системе, в каждый момент времени формируются моменты наступления «ближайшего события в будущем». Затем модельное время продвигается до момента наступления ближайшего из всех возможных событий. В зависимости от того, какое событие оказалось ближайшим, выполняются те или иные действия. Если ближайшим событием является поступление заявки в систему, то выполняются действия, связанные с занятием прибора при условии, что он свободен, и занесение заявки в очередь, если прибор занят. Если же ближайшим событием является завершение обслуживания заявки в приборе, то выполняются действия, связанные с освобождением прибора и выбором на обслуживание новой заявки из очереди, если последняя не пуста. Затем формируется новый момент наступления этого же события. На третьей диаграмме «Модельное (реальное) время» (рис.6.2) продвижение времени в соответствии с этим принципом показано в виде стрелок.

Для того чтобы обеспечить правильную временную последовательность событий в имитационной модели, используются **системные часы**, хранящие значение текущего *модельного времени*. Изменение значения модельного времени осуществляется в соответствии с принципом «пересчёта времени до ближайшего события». Например, если текущее значение модельного времени равно 25, а очередные события должны наступить в моменты времени 31, 44 и 56, то значение модельного времени увеличивается сразу на 6 единиц и «продвигается» до значения 31. Отметим, что единицы времени в модели не обязательно должны быть конкретными единицами времени, такими как секунда или час. Основной единицей

времени в модели можно выбрать любую единицу, которая позволит получить необходимую точность моделирования. Важно помнить, единицы времени выбираются исходя из требований пользователя к точности моделирования. Какая бы единица ни была выбрана, например миллисекунда или одна десятая часа, она должна неизменно использоваться во всей модели.

Кроме рассмотренной службы времени в имитационной модели необходимо реализовать процедуры, связанные с формированием потоков заявок и имитацией обслуживания, с организацией очередей заявок, с организацией сбора и статистической обработки результатов моделирования.

Таким образом, имитационное моделирование дискретных систем со стохастическим характером функционирования, таких как системы и сети массового обслуживания, предполагает использование ряда типовых процедур, обеспечивающих реализацию соответствующих имитационных моделей. К таким процедурам, в первую очередь, относятся следующие процедуры:

- выработка (генерирование) случайных величин:
  - равномерно распределенных;
  - с заданным законом распределения;
- формирование потоков заявок и имитация обслуживания;
- организация очередей заявок;
- организация службы времени;
- сбор и статистическая обработка результатов моделирования
- 

### 3.1.2 Сети Петри

Часто аналитики в задачах моделирования и анализа сложных параллельных и асинхронных систем, обращаются к формальным системам, основанным на использовании математического аппарата сетей Петри. Формальная часть теории сетей Петри, основанная в начале 60-х годов немецким математиком Карлом А. Петри, в настоящее время содержит большое количество моделей, методов и средств анализа, имеющих обширное количество приложений практически во всех отраслях вычислительной техники [39].

Прикладная теория сетей Петри связана главным образом с применением сетей Петри к моделированию систем, их анализу и получающимся в результате этого глубоким проникновением в моделируемые системы.

Моделирование в сетях Петри осуществляется на событийном уровне. Определяются, какие действия происходят в системе, какие состояния предшествовали этим действиям и какие состояния примет система после выполнения действия. Выполнения событийной модели в сетях Петри описывает поведение системы. Анализ результатов выполнения может сказать о том, в каких состояниях пребывала или не пребывала система, какие состояния в принципе не достижимы. Однако, такой анализ не дает числовых характеристик, определяющих состояние системы. Развитие теории сетей Петри привело к появлению, так называемых, «цветных» или «раскрашенных» сетей Петри. Понятие цветности в них тесно связано с понятиями переменных, типов данных, условий и других конструкций, более приближенных к языкам программирования.

Таким образом, структура сети Петри задается ориентированным двудольным мультиграфом, в котором одно множество вершин состоит из позиций, а другое множество – из переходов [9, 11, 15, 19, 21], причем множество вершин этого графа

разбивается на два подмножества и не существует дуги, соединяющей две вершины из одного подмножества.

Итак, сеть Петри – это набор

$$N = (T, P, A), T \cap P = \emptyset,$$

где  $T = \{t_1, t_2, \dots, t_n\}$  - подмножество вершин, называемых переходами;

$P = \{p_1, p_2, \dots, p_n\}$  - подмножество вершин, называемых позициями (местами);

$A \subseteq (T \times P) \cup (P \times T)$  - множество ориентированных дуг.

В сетях Петри вводятся объекты двух типов: динамические – изображаются метками (маркерами) внутри позиций и статические – им соответствуют вершины сети Петри.




Распределение маркеров по позициям называют маркировкой. Маркеры могут перемещаться в сети. Каждое изменение маркировки называют событием, причем каждое событие связано с определенным переходом. Считается, что события происходят мгновенно и одновременно при выполнении некоторых условий.

Каждому условию в сети Петри соответствует определенная позиция. Совершению события соответствует срабатывание (возбуждение или запуск) перехода, при котором маркеры из входных позиций этого перехода перемещаются в выходные позиции. Последовательность событий образует моделируемый процесс. Перемещаемые по сети маркеры часто называют фишками

Основные элементы сети Петри представлены в табл. 3.2.

Таблица 3.2

Элементы сетей Петри

Название элемента	Изображение элемента
Позиция (P)	
Переход (T)	
Дуга	

Переходы в сети Петри являются событиями, которые изменяют состояния в реальной системе. На рис. 3.1 приведен пример интерпретации сети Петри.

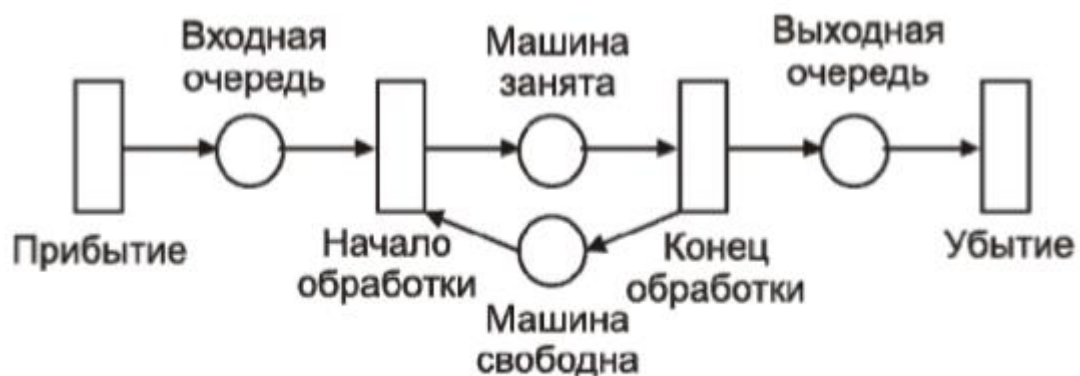


Рис. 3.1. Интерпретация сети Петри

Формальный аппарат сетей Петри предназначен для моделирования систем различного рода и отражает состояния исследуемой системы состоянием сети. Состояние сети Петри определяется ее маркировкой. Количество и распределение фишек сети определяют динамику исследуемой системы. Сеть Петри выполняется посредством запусков переходов в результате удаления фишек из его входных позиций и добавления их в выходные позиции перехода. Последовательность срабатываний переходов полностью определяет поведение сети. Таким образом, сеть Петри описывает структуру системы, ее состояние и поведение [21].

При введении ряда дополнительных правил и условий в алгоритмы моделирования получают различные разновидности сетей Петри. Это необходимо для определения модельного времени, которое позволит моделировать не только последовательность событий, но и их привязку ко времени. В настоящее время выделяют следующие разновидности сетей Петри:

1. Временная сеть Петри – переходы обладают весом, определяющим продолжительность срабатывания (задержку).
2. Стохастическая сеть Петри — задержки являются случайными величинами.
3. Функциональная сеть Петри — задержки определяются как функции некоторых аргументов, например, количества меток в каких либо позициях, состояния некоторых переходов.
4. Цветная (раскрашенная) сеть Петри — метки могут быть различных типов, обозначаемых цветами, тип метки может быть использован как аргумент в функциональных сетях.

Основными свойствами сети Петри являются:

1. Ограниченность – число меток в любой позиции сети не может превысить некоторого значения  $K$ .
2. Безопасность – частный случай ограниченности.
3. Сохраняемость – постоянство загрузки ресурсов.
4. Достижимость – возможность перехода сети из одного заданного состояния (характеризуемого распределением меток) в другое.
5. Живость – возможностью срабатывания любого перехода при функционировании моделируемого объекта.

Среди достоинств аппарата сетей Петри можно указать следующие:

- позволяет моделировать асинхронность и недетерминизм параллельных

независимых событий (в сети Петри могут одновременно и независимо друг от друга сработать несколько переходов), конфликтные взаимодействия между процессами;

- позволяет использовать единые методологические позиции для описания программного обеспечения, аппаратных средств и информационного обмена между системами;

- предоставляет возможность введения любой степени иерархической детализации описываемых программных и аппаратных подсистем модели;

- имеет большую анализирующую мощность, которая позволяет формальными средствами доказывать существование или отсутствие определенных состояний сети Петри.

Однако формальная модель сетей Петри, в силу своей универсальности, имеет ряд недостатков, затрудняющих практическое применение для моделирования сложных систем. К основным таким недостаткам можно отнести следующие:

- высокая трудоемкость анализа сетей большой размерности, а реальные бизнес-процессы предприятия моделируются именно сетями большой размерности;

- описательная мощность сетей Петри недостаточна для содержательного моделирования систем;

- обычные сети Петри не отражают требуемые временные характеристики моделируемой системы;

- фишка сети Петри не представляет собой никакой информации, кроме самого факта ее наличия, поэтому чрезвычайно сложно отразить преобразование информации при срабатывании переходов сети Петри;

- невозможность проведения логических преобразований и, как следствие, – невозможность управления продвижением фишек по сети.

Недостатки сетей Петри не позволяют описывать сложные системы и в настоящее время используются для описания простейших операций. Также эти факторы явились причиной разработки подклассов и расширений сетей Петри, в которых вводятся определенные ограничения на структуру сети, что позволяет использовать более простые алгоритмы для ее анализа либо дополнительные элементы формальной системы, призванные увеличить ее описательную мощность.

Большого внимания заслуживают сети высокого уровня, такие, как раскрашенные сети Петри (Color Petri Net), являющиеся модификацией сетей Петри и отличающиеся хорошо разработанным математическим аппаратом, широко применяемые для самых разнообразных практических целей. Основной причиной высокой эффективности этих формальных моделей является то, что они без потери возможностей формального анализа позволяют исследователю получить значительно более краткие и удобные описания, чем те, которые могут быть сделаны с помощью сетей низкого уровня. В сетях высокого уровня сложность моделей может быть разделена между структурой сети, надписями и описаниями. Это позволяет осуществлять описание значительно более сложных систем и анализировать процессы преобразования данных с помощью общепринятых математических выражений вместо сложного набора позиций, переходов и дуг. Раскрашенные сети Петри, в отличие от обычных сетей Петри, позволяют описывать структуру системы в виде иерархии диаграмм.

Но у данного аппарата моделирования также не устранен ряд недостатков, которые присущи сетям Петри. К таким недостаткам можно отнести:

- необходимость знания разработчиком специфического языка описания моделей [12];
- отсутствие использования принципов объектно-ориентированного подхода;
- низкую гибкость и трудоемкость описания систем в случае их декомпозиции до уровня некоторых элементарных бизнес-операций.

Раскрашенные сети Петри до сих пор применяются для моделирования сложных систем.

Все недостатки СМО и сетей Петри учтены и устранены разработчиками ПП Arena 7.0. Кроме того, этот программный пакет имеет множество необходимых операторов, законов распределения и других элементов, которые привели к его широкому распространению.

Хотелось бы добавить несколько слов о том, почему Arena 7.0 является программным пакетом. Это связано с тем, что Arena 7.0, кроме основного модуля моделирования и анализа систем, имеет следующие встроенные программные средства:

1. ***Input Analyzer.*** Это средство позволяет анализировать входные данные, определять закономерности входных данных для дальнейшего их использования при моделировании систем.
2. ***Output Analyzer*** Это средство позволяет анализировать выходные данные, полученные в результате проведенных экспериментов с моделью.
3. ***Process Analyzer*** Меняет значения параметров модели, структуру модели, занятость ресурсов, их полезность и т. д., сравнивает альтернативные сценарии и выбирает тот сценарий, который имеет наилучший результат. Сравнивая эти сценарии работы модели, можно определить лучшее решение (но не оптимальное, т. к. нельзя просмотреть все возможные решения, т. е. исследовать полностью область допустимых решений), но все-таки определить лучшее решение таким способом возможно.
4. ***Генератор отчетов.*** Выводит данные по результатам моделирования в виде текстовых данных, графиков, диаграмм.
5. ***Visio Process Analyzer.***
6. ***OptQuest.*** Является инструментом оптимизации задач, предназначен и специально настроен для анализа результатов моделирования, выполненного с помощью пакета Arena.

Система имитационного моделирования Arena – основной программный продукт Systems Modeling. Корпорация Systems Modeling была основана в 1982 г. Деннисом Педгеном, автором SIMAN – первого промышленно-ориентированного общецелевого языка имитационного моделирования. В настоящее время область деятельности Systems Modeling включает в себя имитационное моделирование и разработку технологического программного обеспечения [30, 32, 34].

Система Arena позволяет моделировать виды деятельности, представленные на

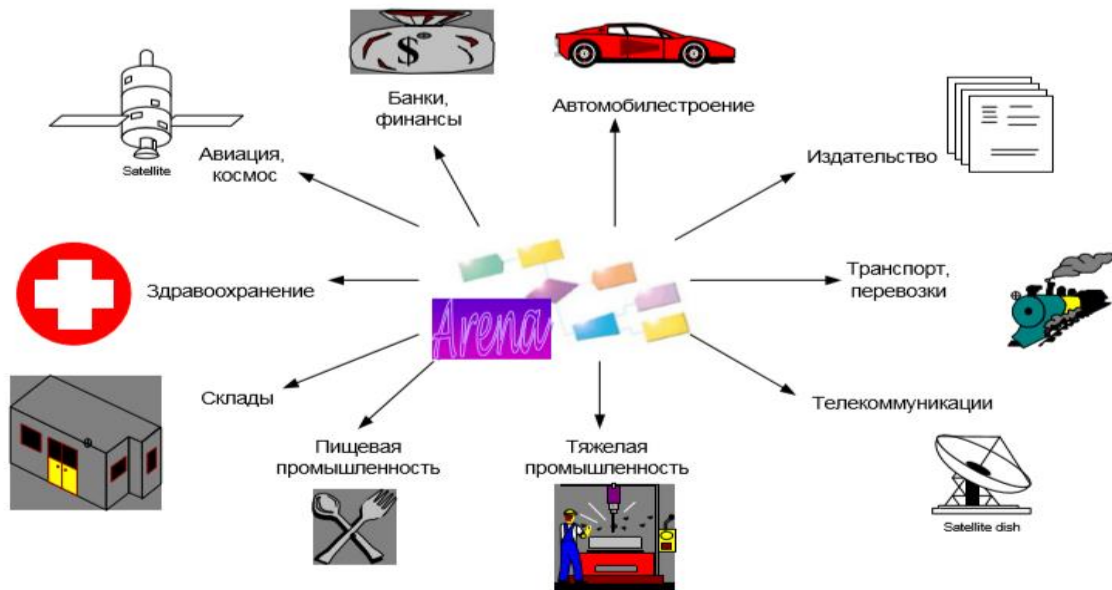


Рис. 3.2. Области применения Arena

рис. 3.2.

С помощью Arena можно достичь основных целей моделирования сложных систем:

- понять, как устроен исследуемый объект: какова его структура, основные свойства, законы развития и взаимодействие с окружающей средой;
- выявить «узкие места» в материальных, информационных и других потоках;
- выделить переменные, наиболее важные для успешного функционирования моделируемой системы, и проанализировать имеющиеся между ними связи;
- научиться управлять системой, определять наилучшие способы управления при заданных целях и критериях;
- прогнозировать прямые и косвенные последствия реализации заданных форм и способов воздействия на систему.

Данные по моделированию в ПП Arena, модулям, их свойствам, можно найти в [30, 32, 33, 34].

#### Список основной литературы

1. *Есипов Б.А.* Методы исследования операций: Учебное пособие. – 2-е изд., испр. и доп. – СПб.:Издательство «Лань», 2013. – 330 с.: ил. – (Учебники для вузов.
2. *Таха Хэмди, А.* Введение в исследование операций. — 6-е изд. — М. : Изд. дом «Вильямс», 2001. — 912 с.
3. *Вентцель, Е. С.* Исследование операций. — М. : Советское радио, 1972. — 550 с.
4. *Алиев Т.И.* Основы моделирования дискретных систем. – СПб: СПбГУ ИТМО, 2009. – 363 с.
5. **Лабораторные работы по курсу «Системы и сети массового обслуживания»:** [Электронный ресурс]: метод. указания / сост. *Б.А. Есипов* – Самара: Изд-во Самарского университета, 2017. – 36 с. : ил. Электрон. текстовые и граф. дан. ( Кбайт).- 1 эл. опт. диск (CD-ROM)



(Алиев)

1. Авен О.И., Гурин Н.Н., Коган Я.А. Оценка качества и оптимизация вычислительных систем. – М.: Наука, 1982. – 464 с.
2. Алиев Т.И. Математические методы теории вычислительных систем. – Л.: ЛИТМО, 1979. – 92 с.
3. Алиев Т.И. Исследование методов диспетчеризации в цифровых управляющих системах. Уч. пособие. – Л.: ЛИТМО, 1986. – 82 с.
4. Бражник А.Н. Имитационное моделирование: Возможности GPSS World. – СПб.: Реноме, 2006. – 439 с.
5. Венцель Е.С. Исследование операций: задачи, принципы, методология. – М.: Наука, 1980. – 408 с.
6. Жожикашвили В.А., Вишневский В.М. Сети массового обслуживания. Теория и применение к сетям ЭВМ. – М.: Радио и связь, 1988. – 192 с.: ил.
7. Кельтон В., Лоу А. Имитационное моделирование. Классика CS. 3-е изд. – СПб.: Питер; Киев: Издательская группа BHV, 2004. – 847 с.: ил.
8. Клейнрок Л. Теория массового обслуживания. Пер. с англ. – М.: Машиностроение, 1979.
9. Клейнрок Л. Вычислительные системы с очередями. – М.: Мир, 1979. – 600 с.
10. Липаев В.В., Колин К.К., Серебровский Л.А. Математическое обеспечение управляющих ЭВМ. – М.: Советское радио, 1972. – 528 с.
11. Основы теории вычислительных систем / С.А.Майоров, Г.И.Новиков, Т.И.Алиев, Э.И.Махарев, Б.Д.Тимченко. – М.: Высшая школа, 1978. – 408 с.
12. Рыжиков Ю.И. Теория очередей и управление запасами: Учебник для вузов. – СПб.: Питер, 2001 год. – 384 с.
13. Советов Б.Я., Яковлев С.А. Моделирование систем: Учебник для вузов. – 4-е изд., стер. – М.: Высшая школа, 2005. – 343 с.: ил.
14. Столингс В. Современные компьютерные сети. – СПб.: Питер, 2003. – 783 с.: ил.
15. Феррари Д. Оценка производительности вычислительных систем. – М.: Мир, 1981.
16. Шварц М. Сети ЭВМ. Анализ и проектирование: Пер. с англ./ Под ред. В.А.Жожикашвили. – М.: Радио и связь, 1981. – 336 с.: ил.
17. Шнепс М.А. Системы распределения информации. Методы расчета. – М.: Связь, 1979.
18. Шрайбер Т.Дж. Моделирование на GPSS. – М.: Машиностроение, 1980.

#### **Список использованных источников**

(Замятина)

1. Бахвалов Л. А. Компьютерное моделирование: долгий путь к сияющим вершинам [Электронный ресурс]. – Режим доступа: <http://www.gpss-forum.narod.ru/GPSSmodeling.html>, свободный.
2. Бешенков С. А. Моделирование и формализация: методическое пособие. – М.: Лаборатория базовых знаний, 2002.



3. Большаков А. С. Моделирование в менеджменте: учеб. пособие. – М.: Филинь, 2000.
4. Бусленко Н. П. Моделирование сложных систем. – М.: Наука, 1978.
5. Бычков С. П., Храмов А. А. Разработка моделей в системе моделирования GPSS: учеб. пособие. – М.: МИФИ, 1997.
6. Введение в математическое моделирование: учеб. пособие / под ред. П. В. Трусова. – М.: Интермет инжиниринг, 2000.
7. Вендров А. М. CASE-технологии. Современные методы и средства проектирования информационных систем. – М.: Финансы и статистика, 1998.
8. Волчков С., Балахонова И. Бизнес-моделирование для совершенствования деятельности промышленного предприятия // ЦИТ «Платон» "КомпьютерПресс". 2001. №11.
9. Докукин В. П. Основы математического моделирования: конспект лекций. Санкт-Петербургский ГГИ. – М.: Дело, 2000.
10. Имитационное моделирование производственных систем / под ред. А. А. Вавилова. – М.: Машиностроение, 1983.
11. Калашников В. В., Рачев С. Т. Математические методы построения стохастических моделей обслуживания. – М.: Наука, 1988.
12. Киндлер Е. Языки моделирования. – М.: Энергия, 1985.
13. Клейнен Дж. Статистические методы в имитационном моделировании. – М.: Статистика, 1978.
14. Лоу А. М., Кельтон В. Д. Имитационное моделирование. Классика CS. – 3-е изд. – СПб.: Питер; Киев: Издательская группа BHV, 2004. – 847 с.: ил.
15. Лукасевич И. Я. Имитационное моделирование инвестиционных рисков [Электронный ресурс]. – Режим доступа: [http://www.cfin.ru/finanalysis/imitation\\_model-2-1.shtml](http://www.cfin.ru/finanalysis/imitation_model-2-1.shtml), свободный.
16. Марков А. А. Моделирование информационно-вычислительных процессов: учеб. пособие. – М.: Изд-во МГТУ им. Н. Э.Баумана, 1999.
17. Максимей И. В. Имитационное моделирование на ЭВМ. – М.: Радио и связь, 1988.

- 18.Марка Д. А., МакГоуэн К. Методология структурного анализа и проектирования. – М.: МетаТехнология, 1993.
- 19.Математическое моделирование: методы, описания и исследования сложных систем / под ред. А. А. Самарского. – М.: Наука, 1989.
- 20.Перегудов Ф. И. Основы системного анализа: учебник. – 2-е изд., доп. – Томск: Изд-во НТЛ, 1997.
- 21.Питерсон Дж. Теория сетей Петри и моделирование систем: пер. с англ. – М.: Мир, 1984.
- 22.Рапопорт Б. М. Инжиниринг и моделирование бизнеса. – М: Тандем, 2001.
- 23.Романовский И. В. Исследование операций и статистическое моделирование. – СПб.: СПб. гос. ун-т, 1994.
- 24.Советов Б. Я. Моделирование систем: учебник для вузов. – 3-е изд., перераб. и доп. – М.: Высш. шк., 2001.
- 25.Суворова Н. Информационное моделирование: величины, объекты, алгоритмы. – М.: Лаборатория базовых знаний, 2002.
- 26.Федотова Д. Э., Семенов Ю. Д., Чижик К. Н. CASE-технологии. – М.: Горячая линия – Телеком, 2003.
- 27.Шебеко Ю. А. Имитационное моделирование и ситуационный анализ бизнес-процессов принятия управленческих решений (учебное и практическое пособие). – М.: Диаграмма, 1999.
- 28.Шеннон Р. Ю. Имитационное моделирование систем – искусство и наука: пер. с англ. – М.: Мир, 1978.
- 29.Щепетова С. Е. Динамическое моделирование функционирования предприятия и формирование стратегии его поведения в конкурентной среде: автореф. дис. на соискание ученой степени к.э.н. – М.: Финансовая академия при Правительстве РФ, 2001.
- 30.ARENA Users Guide, Sewickley: Systems Modeling Co., 1996.
- 31.Barjis J., Shishkov B. UML based business systems modeling and simulation. – Proceedings of EuroSim 2001, 2001.
- 32.Giaglis G. M., Paul R.G., Okeefe R. M. Discrete simulation for business simulation. – Berlin: Springer – Verlag, 2003.
- 33.Goldsmann D. A Whirlwind Tour of Computer Simulation Techniques. – [www.hyperionics.com](http://www.hyperionics.com), 2003.
- 34.Hlupic V., Robinson S. Business Process Modeling and Analysis using discrete-event simulation. – Proceedings of the 1998 Winter Simulation Conference, pp.1363–1369.
- 35.Шеер А.-В. Моделирование бизнес-процессов. – М.: Весть-МетаТехнология, 2000.
- 36.Шеер А.-В. Бизнес-процессы. Основные понятия. Теория. Методы. – М.: Весть-МетаТехнология, 1999.

37. Каменнова М. С., Громов А. И., Ферапонтов М. М., Шматалюк А. Е. Моделирование бизнеса. Методология ARIS. – М.: Весть-МетаТехнология, 2001.
38. Розенберг В. Я., Прохоров А. И. Что такое теория массового обслуживания. – М.: Изд-во «Советское радио», 1962.
39. Малышкин В.Э. Основы параллельных вычислений: Учебное пособие / Часть 1. – Новосибирск: Изд-во НГТУ, 1998.