

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

Т.И. Алиев

ОСНОВЫ ПРОЕКТИРОВАНИЯ СИСТЕМ

Учебное пособие

 **УНИВЕРСИТЕТ ИТМО**

Санкт-Петербург

2015

Алиев Т.И. Основы проектирования систем. – СПб: Университет ИТМО, 2015. – 120 с.

В пособии излагаются основные понятия и определения теории сложных систем и рассматриваются принципы проектирования систем с использованием методов математического моделирования. В качестве моделей дискретных систем для решения задач структурного и функционального проектирования используются модели массового обслуживания, применение которых иллюстрируется на примере вычислительных систем и сетей.

Учебное пособие предназначено для студентов, обучающихся по магистерским программам направления «09.01.04 - Информатика и вычислительная техника» и может быть использовано выпускниками (бакалаврами и магистрантами) при написании выпускных квалификационных работ, связанных с проектированием и исследованием дискретных систем, представляемых в виде моделей массового обслуживания.

Рекомендовано к печати Ученым советом факультета компьютерных технологий и управления, 20 октября 2015 года, протокол № 8.



Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2015

©...Алиев Т.И., 2015

Оглавление

Введение	7
1. Общие вопросы проектирования	9
1.1. Основные понятия	9
1.1.1. Задачи и виды проектирования	9
1.1.2. Проектная документация	10
1.2. Этапы проектирования технических систем	10
1.2.1. Разработка технического задания	11
1.2.2. Этап предварительного проектирования	11
1.2.3. Эскизное проектирование	12
1.2.4. Техническое проектирование	13
1.2.5. Состав проектной документации	14
1.2.6. Автоматизация проектирования	15
1.3. Система как объект проектирования	15
1.3.1. Понятие системы	17
1.3.2. Структурная и функциональная организация системы	18
1.3.3. Общие свойства систем	19
1.3.4. Эффективность системы	21
1.3.5. Параметры и характеристики системы	24
1.3.6. Понятие процесса	28
1.3.7. Классификация систем и процессов	29
1.4. Модель как средство проектирования	31
1.4.1. Типы моделей	31
1.4.2. Основные требования к модели	32
1.4.3. Классификация математических моделей	35
1.4.4. Параметризация моделей	36
2. Типовые задачи и методы проектирования	37
2.1. Типовые задачи проектирования	37
2.1.1. Разработка математической модели	38
2.1.2. Разработка метода исследования	38
2.1.3. Анализ свойств системы	39
2.1.4. Синтез системы и разработка проекта	39
2.1.5. Детальный анализ спроектированной системы	40
2.1.6. Корректировка модели	40
2.2. Методы проектирования	41
2.2.1. Аналитические методы	41
2.2.2. Статистические (имитационные) методы	42
2.2.3. Комбинированный подход	43
2.3. Принципы проектирования систем	44
2.3.1. Системный подход и системотехническое проектирование	44
2.3.2. Принцип иерархического многоуровневого моделирования	45

2.3.3. Принцип множественности моделей	48
2.4. Этапы проектирования систем	49
2.4.1. Формулировка целей проектирования и требований к разрабатываемой системе	49
2.4.2. Разработка концептуальной модели проектируемой системы	49
2.4.3. Разработка и параметризация математических моделей.....	51
2.4.4. Выбор или разработка методов и средств проектирования	51
2.4.5. Проверка адекватности модели.....	53
2.4.6. Проведение экспериментов на модели и анализ характеристик системы	55
2.4.7. Решение задачи синтеза	56
2.4.8. Детальный анализ спроектированной системы.....	57
3. Математические модели дискретных систем.....	58
3.1. Базовые модели	58
3.1.1. Параметры потока заявок	60
3.1.2. Параметры обслуживания заявок	62
3.1.3. Стратегии управления потоками заявок	62
3.1.4. Режимы функционирования базовых моделей.....	66
3.1.5. Характеристики систем с однородным потоком заявок.....	66
3.1.6. Характеристики систем с неоднородным потоком заявок.....	69
3.2. Сетевые модели.....	70
3.2.1. Типы сетевых моделей	71
3.2.2. Параметры сетевых моделей	72
3.2.3. Характеристики сетевых моделей.....	73
3.2.4. Эквивалентные и толерантные преобразования сетевых моделей	74
4. Модели вычислительных систем.....	76
4.1. Принципы структурно-функциональной организации вычислительных систем	76
4.1.1. Основные понятия теории вычислительных систем	76
4.1.2. Структурная организация вычислительных систем	78
4.1.3. Функциональная организация вычислительных систем	80
4.2. Параметры вычислительных систем.....	84
4.3. Характеристики вычислительных систем	90
4.3.1. Характеристики производительности	92
4.3.2. Характеристики оперативности	93
4.3.3. Характеристики надежности	93
4.3.4. Экономические характеристики	94
4.3.5. Локальные характеристики вычислительных сетей	94
4.4. Базовые модели вычислительных систем	95
4.4.1. Модель серверной обработки.....	95
4.4.2. Модель процессорной обработки	96
4.4.3. Модели многомашинной и многопроцессорной обработки	97

4.4.4. Модель процессорной обработки с неоднородной нагрузкой.....	99
4.4.5. Модели каналов связи	100
4.4.6. Модели узлов связи	104
4.5. Сетевые модели вычислительных систем	106
4.5.1. Модель процессорной обработки в многотерминальной ВС	106
4.5.2. Модель ВС с внешними устройствами.....	108
4.5.3. Модель звена передачи данных	112
4.5.4. Модель сети передачи данных	113
5. Вопросы для самопроверки	114
Список литературы.....	121

Введение

При проектировании больших и сложных систем возникают проблемы, связанные не только со свойствами их составных частей – элементов и подсистем, но также и с закономерностями функционирования объекта в целом, рассматриваемые как общесистемные проблемы. Следствием этого является необходимость решения широкого круга специфических задач, к которым относятся: определение общей структуры системы и требований к элементам и подсистемам, организация взаимодействия между ними, выбор оптимальных режимов функционирования, оптимальное управление протекающими в системе процессами, учёт влияния внешней среды и т.п. По мере усложнения систем всё более значимыми становятся общесистемные вопросы, которые рассматриваются в рамках таких дисциплин, как теория сложных систем, системотехника, моделирование.

Проектирование сложных систем обычно предполагает две стадии:

- макропроектирование или внешнее проектирование, в процессе которого решаются структурно-функциональные задачи в целом;
- микропроектирование или внутреннее проектирование, связанное с разработкой элементов системы как физических единиц оборудования.

Методы внешнего проектирования сложных систем базируются на системном подходе к исследованию процессов функционирования, протекающих в системах.

Внешнее проектирование начинается с формулировки проблемы, которая включает в себя:

- определение целей создания системы и круга решаемых ею задач;
- оценка действующих на систему факторов и определение их характеристик;
- выбор показателей и критерия эффективности системы.

В качестве показателей эффективности выбираются числовые характеристики, оценивающие степень соответствия системы своему назначению. Например:

- для системы управления посадкой самолёта показателем эффективности может служить вероятность успешной посадки;
- для автоматической телефонной станции — среднее время ожидания соединения с абонентом;
- для производственного процесса — среднее количество выпускаемых за смену изделий и т. д.

Результаты проведённых исследований используют для обоснования технического задания на разработку системы.

В соответствии с техническим заданием могут быть определены один или несколько вариантов проектируемой системы, которые заслуживают дальнейшего подробного исследования. Анализ вариантов системы, выполняемый на этапе системного анализа, проводится по результатам математического моделирования с применением аналитических и имитационных методов. Результаты моделирования позволяют определить

значения показателей эффективности системы, обосновать её оптимальную структуру и составить рекомендации по совершенствованию исследуемых вариантов. Математическое моделирование является мощным и эффективным инструментом исследования сложных объектов, систем и процессов в различных областях. Многообразие процессов, протекающих в исследуемых системах и объектах, обуславливает и многообразие математических методов и средств, используемых в процессе проектирования.

Большинство технических систем, в том числе вычислительные системы и сети, описываются в терминах дискретных случайных процессов с использованием вероятностных методов. При этом широкое применение находят математические модели, отражающие структурно-функциональную организацию исследуемых систем, построенные на основе моделей теории массового обслуживания, анализ которых может проводиться аналитическими и статистическими методами. В качестве аналитических методов используются вероятностные методы теории массового обслуживания, теории случайных процессов, в качестве статистических – методы имитационного моделирования.

Пособие содержит основные понятия и определения теории проектирования и затрагивает общие вопросы теории сложных систем, принципы математического моделирования. В качестве моделей, на основе которых могут решаться задачи структурного и функционального проектирования, рассматриваются простейшие модели дискретных систем, используемые, в частности, при проектировании вычислительных систем и компьютерных сетей.

Представленный в конце пособия **список литературы** не претендует на полноту и содержит ограниченный перечень литературных источников, которые в той или иной мере использовались при написании пособия. Этот перечень включает источники, которые условно можно разбить на две группы:

- теоретические основы проектирования и математического моделирования дискретных систем [1-3];
- математические модели вычислительных систем и сетей для решения простейших задач проектирования [3-9].

Учебное пособие предназначено для студентов, обучающихся по магистерским программам направления «Информатика и вычислительная техника» и может быть использовано выпускниками (бакалаврами и магистрантами) при написании выпускных квалификационных работ, связанных с проектированием и исследованием систем, представляемых в виде моделей массового обслуживания. Примерами таких систем могут служить компьютерные сети или их фрагменты – вычислительная система (сервер), узел (маршрутизатор, коммутатор) или канал передачи данных.

1. Общие вопросы проектирования

При изложении любой научно-технической дисциплины одним из важных вопросов является формирование терминологии, служащей в дальнейшем фундаментом для изучения теоретических и практических аспектов данной дисциплины. Целью формирования терминологической основы является установление однозначного соответствия между используемым термином и вкладываемым в него смысловым содержанием.

Ниже определены базовые понятия и термины теории систем, а также изложены принципы, задачи, методы и этапы проектирования с использованием математических моделей на примере систем со стохастическим характером функционирования, рассматриваемых в дальнейшем в качестве объектов проектирования.

1.1. Основные понятия

Проектирование (от лат. *projectus*, буквально – брошенный вперед) – процесс создания проекта – прототипа новой системы (или её вариантов), удовлетворяющей предъявляемым к ней требованиям.

1.1.1. Задачи и виды проектирования

Задачи проектирования решаются как в процессе разработки новой системы, так и в процессе модернизации или реконструкции системы.

Модернизация (от англ. *modern* – современный, новый, усовершенствованный) – обновление системы в соответствии с новыми требованиями и нормами путем замены или добавления нового оборудования, расширяющего, например, функциональные возможности системы и улучшающего ее показатели эффективности. Например, модернизация локальной компьютерной сети некоторой организации, направленная на повышение качества обслуживания пользователей, может заключаться в замене сервера сети (маршрутизатора, коммутатора или канала связи), являющегося узким местом, на более мощный сервер или в добавлении второго сервера, позволяющего разгрузить исходный сервер.

Реконструкция – (от *re-* и лат. *constructio* – построение) – коренное изменение (перестройка) системы с целью усовершенствования, направленное на повышение качества функционирования системы. Например, для той же локальной компьютерной сети реконструкция может означать существенное изменение как состава оборудования и топологии сети (изменение структурной организации), так функциональной организации (метода доступа к общим ресурсам, алгоритма маршрутизации и т.п.).

Существуют различные **виды проектирования**, в том числе:

- инженерное проектирование – проектирование инженерных систем (электросетей, газопроводов и т.п.);
- промышленное проектирование (промышленных объектов);
- техническое проектирование (технических систем и устройств);
- проектирование программных средств и информационных систем;

- другие виды проектирования.

Ниже основное внимание уделяется вопросам проектирования технических систем, к которым относятся, в том числе, вычислительные системы, комплексы и компьютерные сети.

1.1.2. Проектная документация

Техническое проектирование различного рода систем состоит в разработке документации, предназначенной для создания новых систем.

Проект технической системы – это комплект технической (проектной) документации, предназначенной для создания новой системы, ее эксплуатации, ремонта и утилизации.

Процесс проектирования часто включает в себя и **конструирование** – процесс создания материального образа разрабатываемой системы в виде макета (прототипа) или графических изображений (чертежей, эскизов, компьютерных моделей).

Прототип (от греч. *protos* – первый и *typos* – отпечаток, оттиск) – прообраз, образец, оригинал.

Термин «прототип» трактуется по-разному в зависимости от предметной области. В области технических систем прототипом является работающая модель или макет, опытный образец устройства или системы.

Состав проектной документации регламентирован стандартами, в том числе, такими как:

- единая система конструкторской документации (ЕСКД) – стандарт, устанавливающий правила и способы разработки, оформления и обращения технической документации;

- единая система программной документации (ЕСПД) – стандарт, устанавливающий правила и способы разработки, оформления и обращения документации по программному обеспечению.

Комплект проектной документации и материалов, содержащих результаты проектирования, включает в себя:

- проектное задание или эскизный проект,
- технический проект,
- рабочий проект.

В процессе проектирования выполняются технические и экономические расчёты, разрабатываются схемы, графики, пояснительные записки, сметы, калькуляции и описания.

При разработке новых технических систем проводятся научно-исследовательские работы, связанные с разработкой новых и обоснованием предлагаемых технических решений.

1.2. Этапы проектирования технических систем

Проектирование технических систем включает в себя следующие этапы:

- 1) разработка технического задания;
- 2) предварительное проектирование;
- 3) эскизное проектирование;

4) техническое проектирование.
Рассмотрим кратко содержание этих этапов.

1.2.1. Разработка технического задания

Техническое задание (ТЗ) составляется разработчиком (проектировщиком) на основе исходных данных, предоставленных заказчиком, содержит основные технические требования к создаваемой системе и служит основанием для проектирования.

Техническое задание содержит:

- назначение системы;
- область применения проектируемой системы;
- технические требования к технико-экономическим показателям разрабатываемой системы, формулируемые в виде ограничений, налагаемых на показатели эффективности;
- условия эксплуатации (режим и продолжительность эксплуатации, внешние воздействия и т.д.);
- сроки и стоимость разработки;
- возможные особые условия производства и эксплуатации;
- другие дополнительные сведения, которые оказывают влияние на результаты проектирования системы.

Одной из важных составляющих технических требований является перечень (номенклатура) показателей, характеризующих технический уровень разрабатываемой системы, а также требования, предъявляемые к ним в виде ограничений, налагаемых на их численные значения. На основе этих показателей в процессе проектирования формируются критерии эффективности, используемые на различных этапах проектирования.

Обоснование и измерение этих показателей, а также разработка методов количественной оценки качества продукции (в том числе технических систем) реализуется в научной дисциплине, называемой *квалиметрией*.

К основным задачам квалиметрии относятся:

- обоснование номенклатуры показателей качества;
- разработка методов определения показателей качества и их оптимизации;
- разработка принципов построения обобщённых показателей качества (критериев эффективности) и обоснование условий их использования в задачах стандартизации и управления качеством.

В качестве математического аппарата в квалиметрии используются методы линейного, нелинейного и динамического программирования, теория оптимального управления, теория массового обслуживания и т.д.

От качества ТЗ существенно зависят результаты проектирования.

1.2.2. Этап предварительного проектирования

Предварительное проектирование проводится в рамках научно-исследовательской работы (НИР). Остальные этапы относятся к опытно-конструкторской разработке (ОКР).

Результаты предварительного проектирования оформляются в виде технических предложений (аванпроекта), представляющих собой технические и технико-экономические обоснования целесообразности разработки системы, удовлетворяющей требованиям ТЗ, а также выявление возможных вариантов реализации системы и их сравнительный анализ. Технические предложения формируются по результатам выполнения НИР в виде научно-технического отчета, содержащего выводы о новых принципах построения системы, научно обоснованный подход к реализации этих принципов, анализ проведенных экспериментов, а также математически обоснованные варианты построения системы. Результаты НИР могут оказаться отрицательными и свидетельствовать о невозможности реализовать систему, удовлетворяющую заданным в ТЗ требованиям.

На этом этапе выполняется:

- обзор и сравнительный анализ существующих вариантов построения системы с целью определения достоинств и недостатков каждого из них;
- выбор возможных вариантов структурно-функциональной организации разрабатываемой системы или синтез оптимального варианта на основе сформулированного критерия эффективности и результатах математического моделирования;
- в случае нескольких вариантов построения системы – их сравнительный анализ и выбор из них наилучшего варианта на основе сравнения основных характеристик этих вариантов или выбранного критерия эффективности;
- в случае синтеза оптимального варианта на основе математического моделирования – выбор метода проектирования, включая разработку математической модели и формулирование задачи оптимизации.

1.2.3. Эскизное проектирование

На этапе эскизного проектирования выбирается окончательный вариант разрабатываемой системы, оформляемый в виде эскизного проекта, в который входят:

- пояснительная записка, включающая, в частности, описание методов исследований, результаты математического (если необходимо, то и физического) моделирования и выполненных расчётов;
- эскизная техническая документация на спроектированную систему;
- заключение о соответствии спроектированной системы техническому заданию.

Эскизный проект в виде совокупности документов дает представление об устройстве и принципе функционирования системы, а также о соответствии назначению и основным требованиям, предъявляемым к разрабатываемой системе.

Эскизная документация предназначена для изготовления лабораторных и экспериментальных образцов системы или её отдельных частей, требующих экспериментального исследования. На этапе эскизного проектирования для

таких испытаний может оказаться необходимым создание специальных испытательных стендов.

Эскизный проект направляется заказчику для ознакомления и выдачи замечаний, после устранения которых защищается разработчиком перед заказчиком и утверждается им или возвращается на доработку. При утверждении эскизного проекта в случае необходимости на основании содержащихся в нем результатов может быть скорректировано техническое задание.

1.2.4. Техническое проектирование

Процесс проектирования технических систем является итеративными с многократным повторением этапов с целью уточнения структурно-функциональных параметров и улучшения проекта.

Проектирование новой технической системы – это противоречивая задача: с одной стороны, следует использовать все последние достижения науки и техники в данной области, а с другой – обычно имеются многочисленные ограничения, указанные в техническом задании (стоимость, сроки проектирования и т.п.).

Во многих случаях проектирование новых технических систем предполагает разработку новых технических идей и инженерных решений для построения эффективной системы, удовлетворяющей заданным требованиям. Поставленная цель может быть достигнута путём решения следующих задач:

- определение, к какому классу систем относится проектируемая система и построение обобщенной модели систем такого класса;
- сравнительный анализ свойств разных систем такого класса и выявление их достоинств и недостатков;
- разработка новых вариантов построения системы;
- сравнительный анализ новых вариантов и обоснование эффективности выбранного варианта построения проектируемой системы.

Процесс проектирования новой технической системы и решение сформулированных выше задач могут быть реализованы одним из следующих способов:

- разработка технической системы с требуемыми новыми свойствами на основе комбинаций известных решений;
- поиск нужных идей на основе ассоциаций в окружающем мире, включая живой, т.е. использование решений, существующих в других сферах;
- создание на основе воображения и собственной фантазии идеализированных образцов разрабатываемых систем и поиск путей их реализации.

Перечисленные задачи и способы их реализации в процессе проектирования новых технических систем тесно связаны с творческим процессом – изобретательством.

Изобретения могут иметь различную степень новизны – от усовершенствования известного прототипа до новой идеи и открытия.

Последнее выходит за рамки инженерного творчества и относится к творчеству научному.

Между научным и инженерным творчеством имеется принципиальное различие. В науке – это выявление новых фактов и закономерностей, присущих системам и протекающим в них процессам, а также разработка новых принципов и методов построения систем. В инженерно-технической области – это создание на основе этих закономерностей и методов новых образцов техники. Правда, нередко при этом получаются и новые научные результаты, вплоть до открытий. Наиболее часто это происходит при создании принципиально новых видов систем, что само по себе может рассматриваться как научная деятельность, поскольку такие разработки не могут полностью базироваться на научно обоснованных методах расчетов, проектирования и т.д.

При проектировании технических систем широко используются средства вычислительной техники для расчётов и моделирования, что позволяет сократить сроки и повысить качество проектирования.

Техническое проектирование осуществляется на основе эскизного проекта с учётом замечаний со стороны заказчика и возможных изменений в техническом задании. Результатом технического проектирования является комплект технической документации, в состав которой входят конструкторская, программная, технологическая и эксплуатационная документации с техническими решениями по структурно-функциональной организации разрабатываемой системы.

На заключительном этапе проектирования систем разрабатывается *рабочий проект* – комплект конструкторских документов, предназначенных для изготовления и испытания опытного образца (макета) проектируемой системы.

1.2.5. Состав проектной документации

В состав проектной документации в общем случае входят:

- конструкторская документация;
- программная документация;
- технологическая документация;
- эксплуатационная документация.

Конкретный состав проектной документации зависит от разрабатываемой системы и цели проектирования.

Конструкторская документация содержит:

- различные схемы разрабатываемой системы: структурные, функциональные, электрические и т.д.;
- чертежи: общего вида, отдельных узлов и деталей;
- текстовые документы: технические условия (ТУ) на систему, техническое описание системы и отдельных её частей (подсистем).

В программную документацию в соответствии с ЕСПД входят:

- текст и описание программ;
- описание применения;

- руководство оператора;
- руководство системного программиста и т.д.

Технологическая документация включает:

- технологические инструкции;
- технологические (маршрутные) карты;
- чертежи на техническую оснастку и приспособления.

В эксплуатационную документацию входят:

- руководство по эксплуатации (РЭ), содержащее описание изделия, рекомендации по его использованию, техническому обслуживанию, текущему ремонту, правила хранения и транспортировки и т.п.;

- инструкция по монтажу, пуску и т.п.
- ведомость ЗИП – запасных частей, инструмента и приспособлений.

На этапе технического проектирования осуществляются изготовление и испытания опытных, а затем и серийных образцов изделия с последующей корректировкой документации по результатам этих испытаний.

1.2.6. Автоматизация проектирования

Автоматизация проектирования состоит в применении ЭВМ и специальных программных средств для проектирования систем. Автоматизированное проектирование позволяет исключить субъективизм при принятии решений, повысить точность расчётов, предоставить возможность выбора наилучшего, а в некоторых случаях, оптимального варианта для реализации на основе строгого математического анализа нескольких вариантов проекта с оценкой технических, технологических и экономических характеристик системы, значительно повысить качество проектной документации, существенно сократить сроки проектирования.

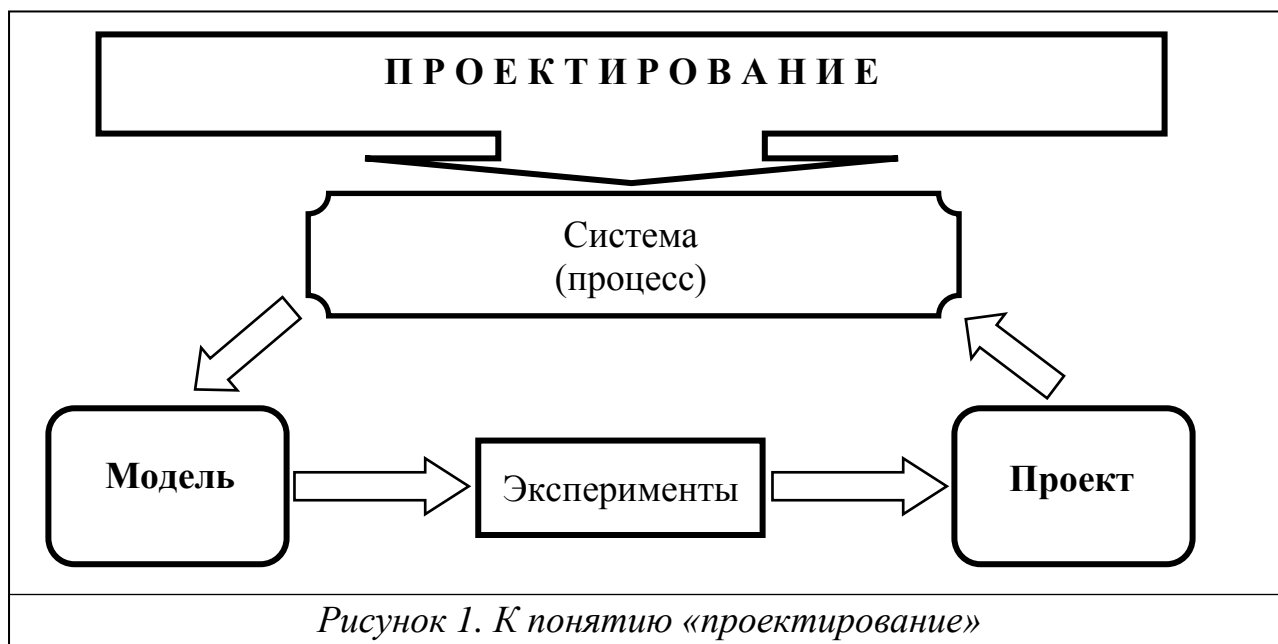
Методы и средства автоматизации проектирования различны и зависят от характера и назначения проектируемой системы. При этом наибольший эффект достигается при автоматизации проектирования больших и сложных технических систем, характеризующихся наличием большого количества разнородных элементов и многочисленными сложными связями между ними. Так, например, при проектировании вычислительных систем и компьютерных сетей с помощью автоматизированной системы проектирования определяются конфигурация (топология) системы, технические параметры входящих в её состав устройств, их структурная и функциональная реализация, рассчитываются показатели производительности, надёжности, экономической эффективности и т.д.

1.3. Система как объект проектирования

Объектами проектирования в технике являются системы и протекающие в них процессы (рисунок 1). В вычислительной технике объектами проектирования являются вычислительные машины, комплексы, системы и сети. В последних, кроме вычислительных процессов, присущих вычислительным машинам и системам, важную роль играют процессы передачи данных. Проектирование таких систем обычно является

многоэтапным и предполагает использование математических методов моделирования.

Моделирование заключается в представлении исследуемой системы (процесса) в виде некоторого объекта, называемого **моделью** (рисунок 1), и проведении **экспериментов** на модели с целью получения информации о системе путем исследования свойств модели, на основе которого формулируются и решаются задачи, связанные с разработкой **проекта** системы, удовлетворяющей перечисленным в техническом задании требованиям.



Моделирование предоставляет возможность решать задачи проектирования систем, прямой эксперимент с которыми трудно выполним, экономически невыгоден или вообще невозможен.

Моделирование как эффективное средство исследования широко применяется в различных областях науки и техники при проектировании: вычислительных систем и компьютерных сетей, систем в авиа- и автомобилестроении, в приборо- и машиностроении, систем телекоммуникаций и т.д. Повсеместное применение моделирования при проектировании технических систем обусловлено всеохватывающим внедрением средств вычислительной техники и появлением специализированных программных средств моделирования в различных областях, используемых 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.

1.3.1. Понятие системы

Под *системой* (от греч. systema – целое, составленное из частей; соединение) будем понимать совокупность взаимосвязанных элементов, объединенных в одно целое для достижения некоторой цели, определяемой назначением системы. Здесь следует обратить внимание на два момента: во-первых, элементы обязательно взаимосвязаны и, во-вторых, система имеет определенное назначение.

Элемент представляет собой минимальный неделимый объект, рассматриваемый как единое целое.

Понятие «система» широко используется в повседневной жизни. Мы говорим «система знаний», «система оценок», «система взглядов» и т.д. Можно заметить, что приведенные термины не вполне соответствуют данному выше определению. И хотя можно попытаться найти в этих «системах» элементы и связи между ними, это всё-таки будет выглядеть несколько искусственно притянутым. Такая ситуация обусловлена тем, что в этих примерах понятие «система» используется в более широком общепринятом смысле. В каждой предметной области может быть введена своя трактовка понятия «система».

Техническая система – это создаваемый человеком на основе достижений науки и техники объект, обладающий структурной и функциональной организацией.

Таким образом, для технических систем будем руководствоваться данным выше определением термина «система», и, говоря о системе, будем иметь в виду, прежде всего, техническую систему.

Сложная (большая) система характеризуется большим числом входящих в его состав элементов и множеством связей между ними.

Комплекс представляет собой совокупность взаимосвязанных систем. Соотношение введенных понятий иллюстрирует рисунок 2.

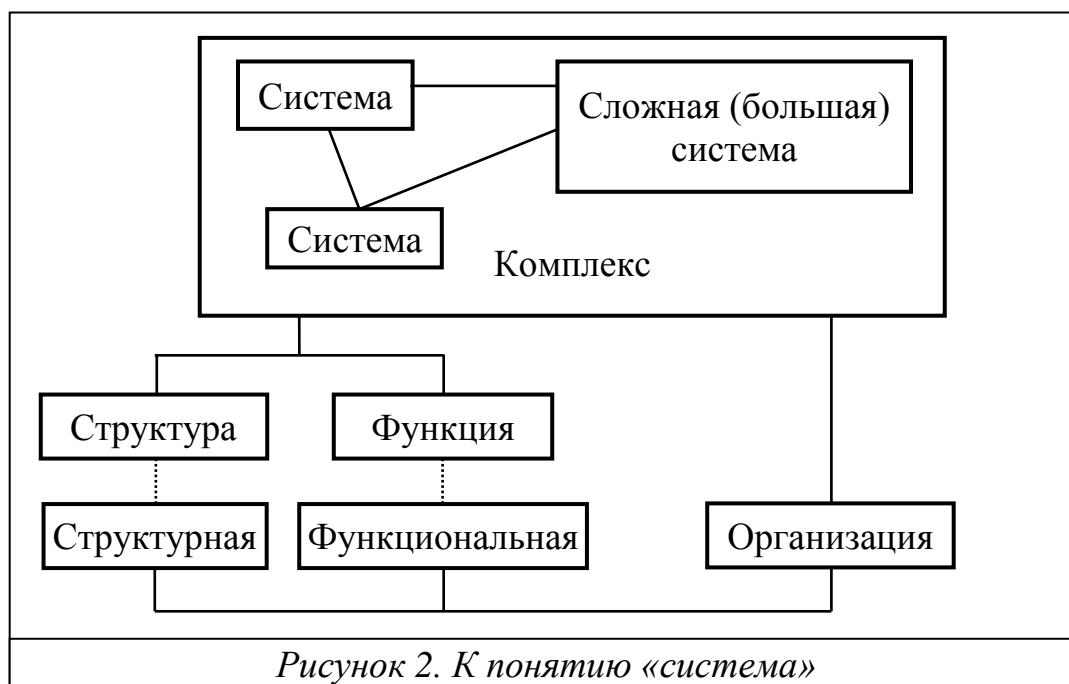


Рисунок 2. К понятию «система»

Элемент, система и комплекс – понятия относительные. Любой элемент может рассматриваться как система, если его расчленишь на более мелкие составляющие – элементы. И наоборот, любой комплекс может рассматриваться как система, если входящие в его состав системы рассматривать как единое целое и трактовать их как элементы. В связи с этим, понятия «система» и «комплекс» часто трактуют как эквивалентные понятия. Например, вычислительную машину можно рассматривать как систему, элементами которой являются центральный процессор, оперативная память, накопители на магнитных дисках, устройства ввода-вывода. В то же время, центральный процессор можно рассматривать как систему, состоящую из таких элементов, как арифметико-логическое устройство, устройство управления, счетчик команд, регистровая память и т.д.

1.3.2. Структурная и функциональная организация системы

Проектирование системы обычно связано с определением её структурной и функциональной организации (рисунок 2).

Под *организацией* системы понимают способ достижения поставленной цели за счет выбора определенной структуры и функции системы и соответственно различают структурную и функциональную организацию системы.

Структурная организация определяется набором элементов и способом их соединения в структуру, обеспечивающую возможность реализации возлагаемых на систему функций.

Функциональная организация определяется способом порождения функций системы, достаточных для достижения поставленной цели. Синонимом термина «функциональная организация» часто служит термин «режим функционирования» системы, определяющий порядок выполнения возложенных на систему функций.

Структура системы задается перечнем и значениями параметров входящих в состав системы элементов и связями между элементами.

Структура технической системы задается, например, в виде перечня и количества устройств, блоков, узлов, их техническими характеристиками и матрицы связей между устройствами.

Структура системы может быть задана следующими способами (рисунок 3):

- *графически* в форме:

- ✓ графа, в котором вершины соответствуют элементам системы, а дуги – связям между ними;
- ✓ схем, широко используемых в инженерных приложениях, в которых элементы обозначаются в виде специальных символов;

- *аналитически* путем задания количества типов элементов, количества и значений параметров (скорость работы, размер и т.п.) элементов каждого типа, а также матрицы связей (инцидентности), определяющей взаимосвязь элементов.

Функция системы представляет собой правило достижения поставленной цели, описывающее поведение системы и направленное на получение результатов, предписанных назначением системы.

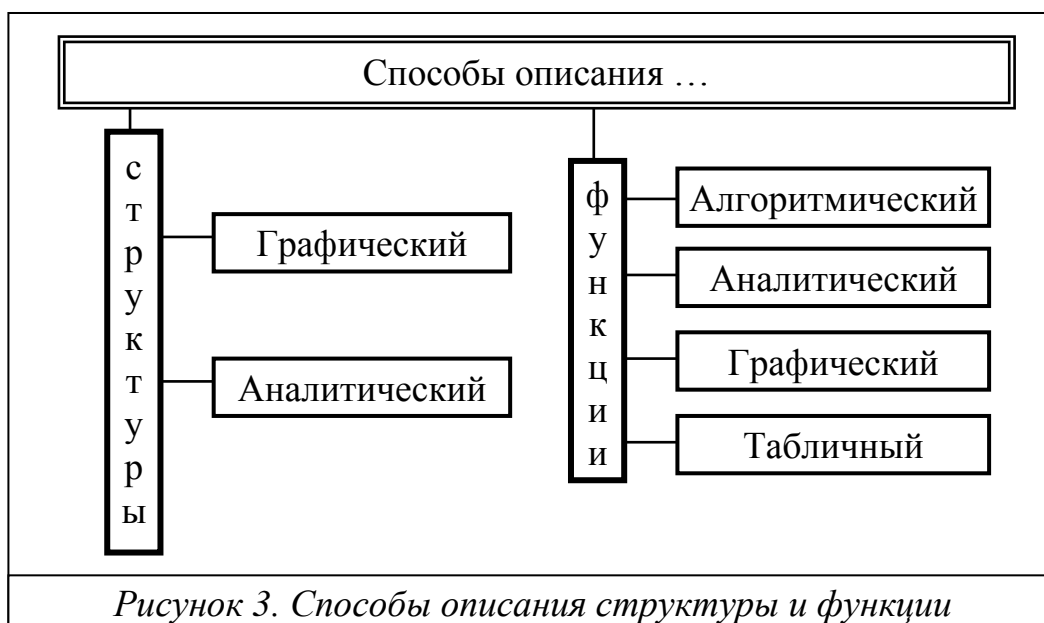
Функция технической системы определяется её назначением и может быть представлена:

- **алгоритмически** в виде словесного описания, содержащего последовательность шагов, которые должна выполнять система для достижения поставленной цели;

- **аналитически** в виде математических зависимостей в терминах некоторого математического аппарата: теории множеств, теории случайных процессов, теории дифференциального или интегрального исчисления и т.п.;

- **графически** в виде временных диаграмм или графических зависимостей;

- **таблично** – в виде различных таблиц, отражающих основные функциональные зависимости, например, в виде таблиц булевых функций, автоматных таблиц функций переходов и выходов и т.п.



Функциональная организация реализуется безотносительно к необходимым для этого средствам (элементам), в то время как структурная организация определяется функцией, возлагаемой на систему.

1.3.3. Общие свойства систем

Любым сложным системам присущи **общие (фундаментальные) свойства**, диктующие необходимость применения системного подхода при их проектировании и исследовании методами математического моделирования. Такими свойствами являются (рисунок 4):

- **целостность**, означающая, что система рассматривается как единое целое, состоящее из взаимодействующих элементов, возможно неоднородных, но одновременно совместимых;

- **связность** – наличие существенных устойчивых связей между элементами и/или их свойствами, причем с системных позиций значение имеют

не любые, а лишь существенные связи, которые определяют *интегративные* свойства системы;

- **организованность** – наличие определенной структурной и функциональной организации, обеспечивающей достижение поставленной цели;

- **интегративность** – наличие качеств, присущих системе в целом, но не свойственных ни одному из ее элементов в отдельности.

Таким образом, можно сделать следующие важные выводы:

- система не есть простая совокупности элементов;
- расчлняя систему на отдельные части и изучая каждую из них в отдельности, нельзя познать все свойства и закономерности, присущие системе в целом.

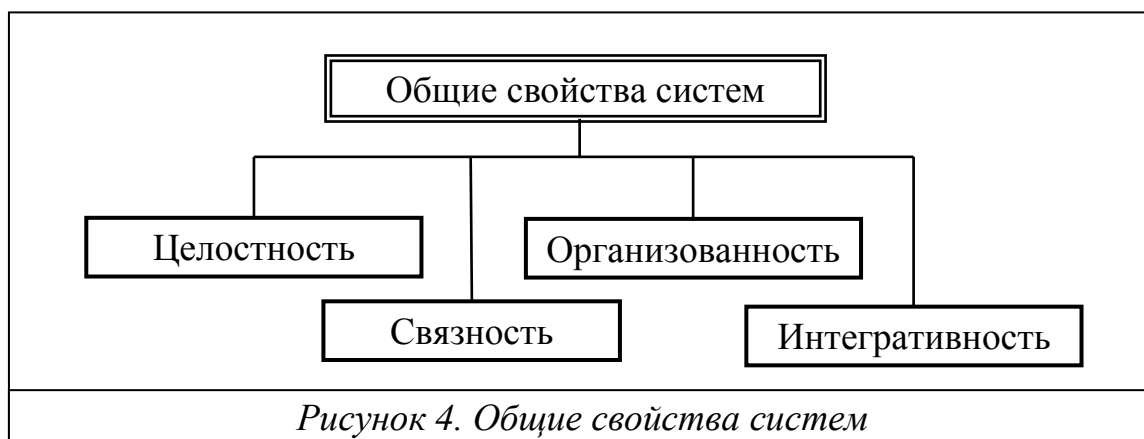


Рисунок 4. Общие свойства систем

С учётом изложенного попытаемся ответить на следующий вопрос: «можно ли персональный компьютер рассматривать как систему, элементами которого являются системный блок и связанные с ним внешние устройства – монитор, принтер и сканер?».

Если воспользоваться определением системы, как совокупности взаимосвязанных элементов, то вроде бы компьютер с внешними устройствами можно считать системой. Однако следует обратить внимание на вторую часть определения понятия «система», где сказано, что элементы, объединенные в одно целое должны обеспечивать достижение цели, определяемой назначением системы. Это означает, что система, кроме структурной организации в виде совокупности взаимосвязанных элементов, должна обладать и функциональной организацией, то есть в ней должны протекать некоторые процессы во времени, изменяющие состояние системы. С этих позиций неработающий компьютер не может трактоваться как система. В то же время, если в этом компьютере выполняется некоторая задача, его можно рассматривать как систему, обладающую структурной и функциональной организацией. Однако и здесь имеются некоторые нюансы, которые следует учитывать при выявлении соответствия рассматриваемого объекта введенному понятию «система».

Вспомним, что система должна обладать такими фундаментальными свойствами, как целостность, связность, организованность и интегративность. Наличие этих свойств позволяет рассматривать систему как единое целое и применять для её исследования системный подход. Особенно важным является

последнее свойство – интегративность, свидетельствующее о том, что невозможно полностью познать систему, анализируя только свойства её элементов. Другими словами, система может обладать свойствами, которые не присущи ни одному из входящих в её состав элементов.

1.3.4. Эффективность системы

Степень соответствия системы своему назначению называется *эффективностью*.

Процесс проектирования системы обычно предполагает решение двух взаимосвязанных задач:

- задач анализа, связанных с оценкой эффективности вариантов структурно-функциональной организации системы, задаваемой в виде совокупности показателей эффективности;

- задач синтеза, направленных на построение оптимальных систем или выбор наилучшего варианта структурно-функциональной организации системы в соответствии с выбранным критерием эффективности.

Анализ (от греч. *análysis* — разложение, расчленение) заключается в определении свойств и закономерностей, присущих процессам, протекающим в системе. В процессе анализа на основе сведений о функциях и параметрах элементов, входящих в состав системы, и сведений о структуре системы определяются значения *показателей эффективности*, описывающие свойства, присущие системе в целом.

При проектировании простых систем задачи анализа иногда решаются на основе декомпозиции (разложения) системы и сводятся к изучению свойств отдельных подсистем и элементов. Полученные результаты распространяются на всю систему. Очевидно, что при этом могут быть потеряны и не учтены некоторые важные свойства, присущие системе в целом и обусловленные наличием такого фундаментального свойства сложных систем, как интегративность, поскольку система может обладать свойствами, не присущими ни одному из элементов.

Синтез (от греч. *synthesis* - соединение, сочетание, составление) системы заключается в определении структурной и функциональной организации системы, удовлетворяющей требованиям, предъявляемым к эффективности системы. Синтез служит основой для задач проектирования, и направлен на построение системы, удовлетворяющей наилучшим образом заданным требованиям. При этом задача синтеза может решаться как задача выбора наилучшего варианта из нескольких вариантов построения системы, либо как задача построения оптимальной системы в соответствии с выбранным критерием эффективности.

Эффективность систем обычно оценивается набором *показателей эффективности*.

Показатель эффективности (качества) – мера одного свойства системы. Показатель эффективности всегда имеет количественный смысл.

Показатели эффективности технической системы можно разделить на:

- функциональные, описывающие качество функционирования системы, к которым относятся производительность, оперативность, надежность и т.п.;
- экономические (стоимость системы, затраты на разработку и эксплуатацию и т.д.);
- технологические (трудоемкость разработки и т.п.);
- антропологические (экологичность, безопасность, защищенность и т.п.).

Количество показателей эффективности технических систем во многих случаях может оказаться достаточно большим. Обычно показатели эффективности являются противоречивыми. Это означает, что изменение структурной или функциональной организации системы приводит к улучшению одних показателей и, в то же время, к ухудшению других показателей эффективности, что существенно осложняет выбор наилучшего варианта (способа) структурно-функциональной организации проектируемой системы. Очевидно, что при проектировании системы предпочтительно иметь один показатель эффективности. Такой показатель называется критерием эффективности.

Критерий эффективности является мерой эффективности системы, обобщающей свойства системы в одной оценке – значении критерия эффективности. Если при увеличении эффективности значение критерия возрастает, то критерий называется **прямым**, если же значение критерия уменьшается, то критерий называется **инверсным**.

Критерий эффективности служит для выбора из всех возможных вариантов структурно-функциональной организации системы наилучшего варианта.

Система, которой из всех возможных вариантов построения системы, удовлетворяющих заданным требованиям, соответствует максимальное (минимальное) значение прямого (инверсного) критерия эффективности называется **оптимальной системой**.

Как сказано выше, для описания системы обычно используется множество зачастую противоречивых показателей эффективности, при этом попытка улучшить какой-то один или несколько показателей эффективности за счет изменения параметров структурно-функциональной организации системы обычно приводит к ухудшению остальных показателей эффективности. Например, если мы хотим построить высокопроизводительную и сверхнадежную вычислительную систему, то, очевидно, что ее стоимость окажется чрезвычайно большой. С другой стороны, если задаться целью – построить как можно более дешевую вычислительную систему, то вряд ли ее производительность и надежность окажутся на должном уровне. Таким образом, для решения задачи оптимального синтеза системы целесообразно иметь одну целевую функцию, то есть один критерий эффективности, позволяющий выбрать из множества вариантов построения системы наилучший, а точнее оптимальный вариант, то есть такой, при котором критерий эффективности принимает максимальное (прямой критерий эффективности) или минимальное (инверсный критерий эффективности)

значение. Существует несколько способов построения критерия эффективности при наличии множества показателей эффективности. Это, прежде всего так называемые составные критерии эффективности (аддитивные или мультипликативные), представляющие собой объединение (сумму или произведение) многих показателей эффективности. Однако на практике более широкое распространение получили критерии эффективности с ограничениями, которые строятся по следующему принципу: из множества показателей эффективности один выбирается в качестве критерия эффективности, а на остальные показатели налагаются ограничения.

Следует отметить, что вид критерия эффективности зависит от назначения системы. Если система предназначена для обеспечения высокой надежности, то в качестве критерия эффективности может использоваться один из показателей надежности. Если же система должна иметь высокую производительность, то в качестве критерия эффективности следует использовать показатель производительности системы. Возможна ситуация, когда к проектируемой системе предъявляются требования и высокой производительности и надежности. Тогда в качестве критерия эффективности можно использовать составной критерий эффективности, объединяющий два показателя эффективности – производительность и надежность.

И еще. В соответствии с изложенным выше должно быть ясно, что оптимальная система существует в единственном экземпляре, для которой значение критерия эффективности имеет максимальное (для прямого критерия) или минимальное (для инверсного критерия) значение. Таким образом, фразы типа «более оптимальная система» или «менее оптимальная система» являются некорректными.

При решении традиционных задач оптимизации обычно используется один критерий эффективности, который формируется одним из вышеуказанных способов в зависимости от цели оптимизации и постановки задачи. В то же время, в литературе часто встречается такое понятие как «многокритериальная задача». Означает ли это, что задача оптимального синтеза может решаться с использованием сразу нескольких критериев эффективности?

Действительно, понятие «многокритериальная задача» достаточно широко используется в такой математической дисциплине как «Исследование операций». Задачи, в которых имеется одна целевая функция (один критерий эффективности), принимающая численные значения, относятся к задачам математического (или оптимального) программирования. Им противостоят задачи с несколькими целевыми функциями или с одной целевой функцией, но принимающей векторные значения или значения ещё более сложной природы. Эти задачи называются многокритериальными и решаются путём сведения к задачам с единственной целевой функцией. Многокритериальными задачами являются задачи теории игр, изучающей формальные модели принятия оптимальных решений в условиях конфликта. При этом под конфликтом понимается явление, в котором участвуют различные стороны, наделённые различными интересами, выраженными в виде целевых функций (критериев эффективности), и возможностями выбирать доступные для них действия в

соответствии с этими интересами. В условиях конфликта стремление противника скрыть свои предстоящие действия порождает неопределённость. Поэтому теория игр рассматривается также как теория принятия оптимальных решений в условиях неопределённости.

Часто в литературных источниках вместо понятия «оптимальная система» используется понятие «рациональная система». Естественно, что возникает вопрос: в чём различие между рациональной и оптимальной системой?

«Оптимальная система» означает, что значения параметров структурно-функциональной организации определены в процессе решения математической оптимизационной задачи и являются оптимальными, то есть обеспечивают экстремум выбранного критерия эффективности. На практике может оказаться невозможным построить систему с такими значениями параметров, что может быть обусловлено разными причинами, в том числе, дискретным характером оптимизируемых параметров.

Например, в процессе синтеза некоторой сети передачи данных получены следующие оптимальные значения пропускных способностей трёх каналов связи: 428 кбит/с, 764 кбит/с и 931 кбит/с. Положим, что реальные каналы связи могут иметь пропускные способности в 256 кбит/с, 512 кбит/с и 1024 кбит/с. Очевидно, что в качестве окончательного решения задачи проектирования будут приняты значения 512 кбит/с, 512 кбит/с (или 1024 кбит/с) и 1024 кбит/с. Поскольку эти значения отличаются от оптимальных, спроектированная система не может считаться оптимальной. Такую систему обычно называют «рациональной», имея в виду, что ее параметры близки, но не равны оптимальным значениям.

Другой случай, когда в результате оптимизации получено значение пропускной способности канала 2000 кбит/с, которое существенно превышает максимально допустимое значение в 1024 кбит/с. Очевидно, что в этом случае одно из возможных решений состоит в установке двух каналов с пропускной способностью 1024 кбит/с, что также не будет соответствовать оптимальному варианту.

Иногда под «рациональной системой» подразумевают некоторый вариант её построения, выбранный из нескольких возможных вариантов на основе анализа характеристик функционирования или сравнения значений критерия эффективности. Ясно, что в этом случае вообще речь не идет об оптимизации.

1.3.5. Параметры и характеристики системы

Любая система может быть описана совокупностью (рисунок 5):

- **параметров**, описывающих первичные свойства системы и являющихся исходными данными при решении задач анализа;

- **характеристик**, описывающих вторичные свойства системы и определяемых в процессе решения задач анализа как функция параметров, то есть являющихся вторичными по отношению к параметрам.

Множество параметров технических систем можно разделить на:

- **внутренние**, описывающие структурно-функциональную организацию системы, к которым относятся:

- ✓ **структурные параметры**, описывающие состав и структуру системы;
- ✓ **функциональные параметры**, описывающие функциональную организацию (режим функционирования) системы.
- **внешние**, описывающие взаимодействие системы с внешней по отношению к ней средой, к которым относятся:
 - ✓ **нагрузочные параметры**, описывающие входное воздействие на систему, например частоту и объем используемых ресурсов системы;
 - ✓ **параметры внешней (окружающей) среды**, описывающие обычно неуправляемое воздействие внешней среды на систему, например помехи и т.п.

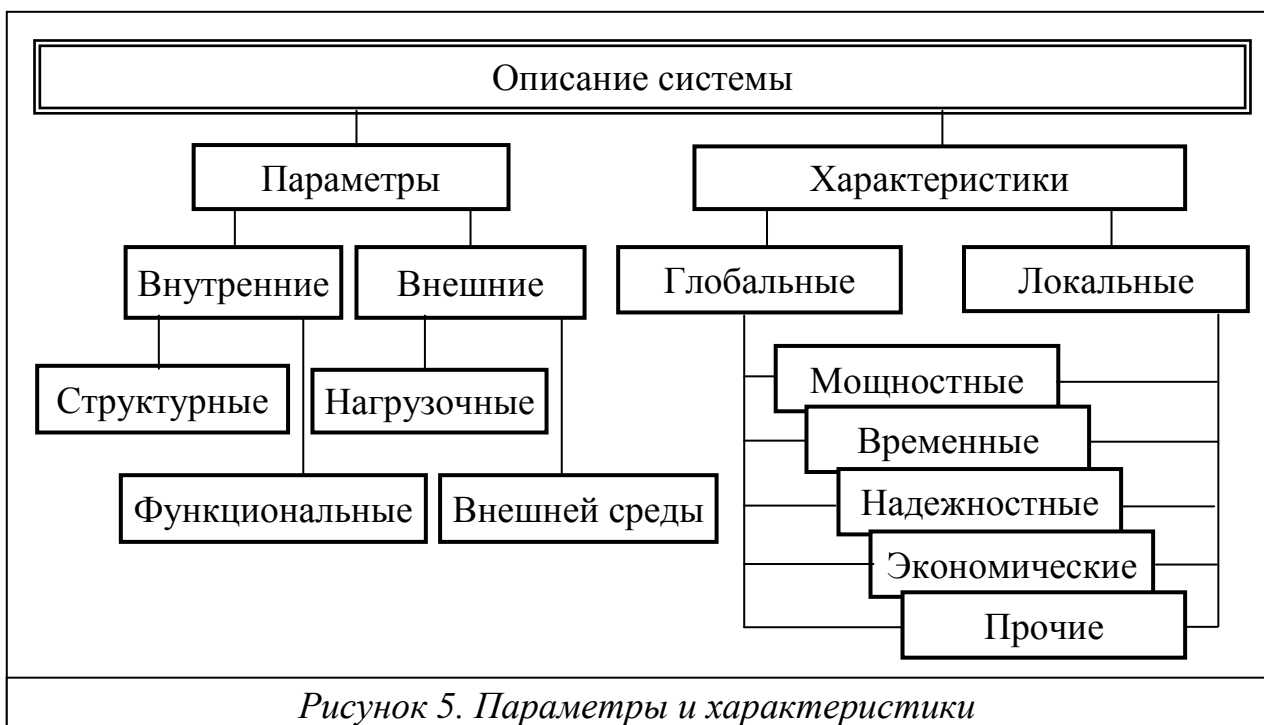


Рисунок 5. Параметры и характеристики

Параметры могут быть:

- детерминированными или случайными;
- управляемыми или неуправляемыми.

Характеристики системы, в отличие от показателей эффективности, могут быть качественными или количественными. Примерами качественных характеристик могут служить:

- функциональность – перечень выполняемых системой функций;
- гибкость – возможность реконфигурирования системы при возникновении специфических ситуаций;
- совместимость – например, возможность использования элементов разных производителей;
- взаимозаменяемость и т.п.

В некоторых случаях качественные характеристики пытаются описать количественно, например функциональность – числом выполняемых системой функций.

И все же, основными при проектировании систем являются количественные характеристики, которые можно объединить в два класса:

- **глобальные**, описывающие эффективность системы в целом;
- **локальные**, описывающие качество функционирования отдельных элементов или частей (подсистем) системы.

Глобальные характеристики технических систем можно разбить на следующие группы:

- **мощностные (характеристики производительности)**, описывающие скоростные качества системы, измеряемые, например, количеством задач, выполняемых вычислительной системой за единицу времени, или количеством пакетов, передаваемых в компьютерной сети за единицу времени;

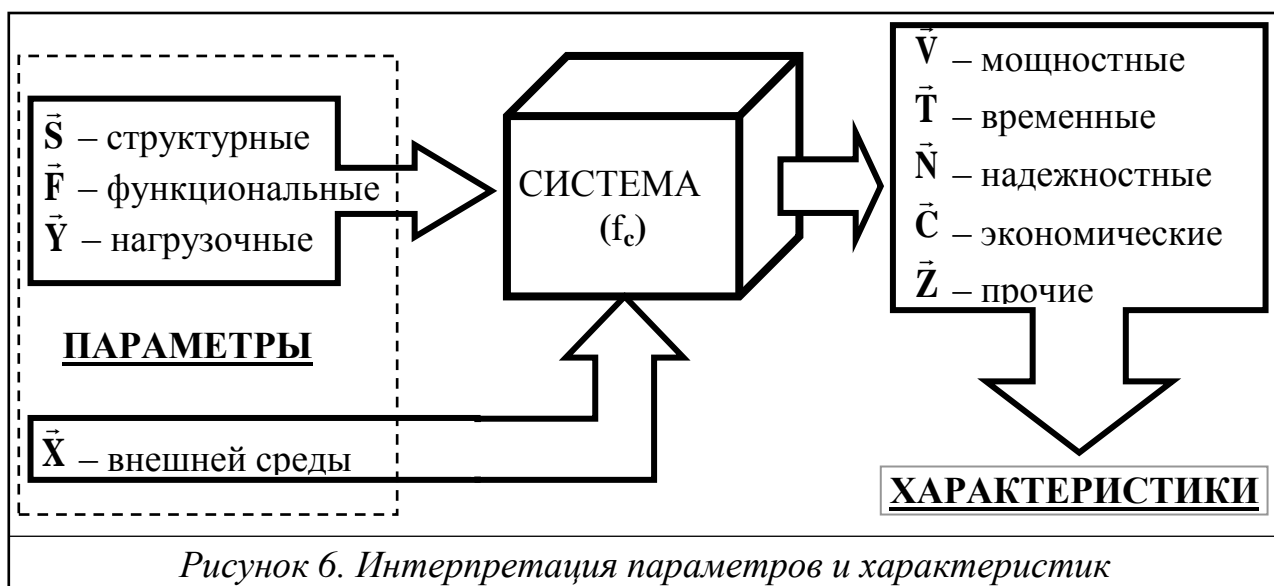
- **временные (характеристики оперативности)**, описывающие временные аспекты функционирования системы, например время выполнения задач в вычислительной системе или время задержки пакетов при передаче в компьютерной сети;

- **надежностные (характеристики надежности)**, описывающие надежность функционирования системы;

- **экономические (стоимостные)** в виде стоимостных показателей, например, стоимость технических и программных средств вычислительной системы, затраты на эксплуатацию компьютерной сети и т.п.;

- **прочие**: массогабаритные, энергопотребления, тепловые и т.п.

Параметры можно интерпретировать как входные величины по отношению к системе, а характеристики – как выходные величины, зависящие от параметров и определяемые в процессе анализа системы (рисунок 6).



В этом случае закон функционирования системы можно представить в виде:

$$\vec{H}(t) = f_c(\vec{S}, \vec{F}, \vec{Y}, \vec{X}, t),$$

где f_c – функция, функционал, логические условия, алгоритм, таблица или словесное описание, определяющее правило (закон) преобразования входных величин (параметров) в выходные величины (характеристики); $\vec{H}(t)$ – вектор характеристик $\vec{H} = \{\vec{V}, \vec{T}, \vec{N}, \vec{C}, \vec{Z}\}$, зависящий от текущего момента времени t ($t \geq 0$); $\vec{S}, \vec{F}, \vec{Y}$ и \vec{X} – векторы параметров структурных, функциональных, нагрузочных и внешней среды соответственно; $\vec{V}, \vec{T}, \vec{N}, \vec{C}$ и \vec{Z} – векторы мощностных (производительности), временных, надежностных, экономических и прочих характеристик соответственно.

В некоторых литературных источниках не делается различия между параметрами и характеристиками. Более того, одни и те же величины могут называться то параметрами, то характеристиками.

Введенные выше определения четко разделяют описывающие систему величины на две группы: «параметры» и «характеристики». Характеристики системы являются функциями параметров, то есть изменение какого-либо параметра приводит к изменению характеристик системы.

В то же время следует помнить, что «параметры» и «характеристики» – понятия относительные. Это можно показать на следующем примере. Если выполняющий некоторые задачи компьютер рассматривается как система, одним из элементов которой является процессор, то производительность (быстродействие) процессора является параметром, изменение которого приведет к изменению такой величины, как время выполнения задачи, которая в данном случае представляет собой характеристику системы. Если же процессор рассматривается как система, состоящая из арифметико-логического устройства, устройства управления, регистровой памяти и т.д., то быстродействие процессора будет являться характеристикой, которая зависит от параметров входящих в ее состав элементов. Можно было бы сказать, что параметры системы в основном описывают элементы системы и их взаимосвязь (как структурную, так и функциональную), а характеристики описывают систему в целом. Однако это будет не совсем корректно, поскольку характеристики могут описывать как систему в целом (глобальные характеристики), так и её отдельные элементы и подсистемы (локальные характеристики).

Еще одним важным моментом при обсуждении термина «характеристика» является выяснение его отличия от термина «показатель эффективности».

Судя по определениям, термины «показатель эффективности» и «характеристика» – близкие понятия. Можно даже предположить, что это одно и то же. Однако между ними существует определенное различие.

Во-первых, как сказано выше, показатель эффективности всегда имеет количественный смысл, т.е. представляется в виде количественной оценки, в то время как характеристика может быть качественной. Так, например, при описании

вычислительных систем и сетей широко используются такие характеристики, как открытость, масштабируемость, гибкость, информационная безопасность и т.п., количественное задание которых либо достаточно условно, либо вообще невозможно.

Во-вторых, множество показателей эффективности при исследовании некоторой системы зависит от её назначения, в то время как характеристики описывают всю совокупность свойств системы. При этом возможно, что некоторые характеристики являются несущественными. Например, если компьютер предназначен для использования в космосе или на борту самолета, то важными показателями эффективности являются его вес и энергопотребление. Если же компьютер предназначен для решения сложных задач моделирования, оптимизации или игровых задач (например, шахматных), требующих большой вычислительной мощности, то более актуальными становятся такие показатели эффективности как производительность, время реакции, а вес и энергопотребление могут вообще не иметь никакого значения.

Таким образом, совокупность всех характеристик системы полностью описывает все свойства системы, в то время как множество показателей эффективности, являясь подмножеством характеристик, отражает только определённые свойства системы, представляющиеся существенными в процессе конкретных исследований.

1.3.6. Понятие процесса

Наличие функциональной организации системы означает, что в ней протекают во времени процессы, изменяющие состояние системы. Так, например, вычислительная система реализует процесс обработки данных, а телекоммуникационная сеть – процесс передачи данных.

Изучение сложных систем удобно проводить в терминах процессов, с которыми связаны такие понятия как «состояние», «переход» и «событие».

Процесс (от лат. *processus* – продвижение) – последовательная смена состояний системы во времени.

Состояние системы задается совокупностью значений переменных, описывающих это состояние. Система находится в некотором состоянии, если она полностью описывается значениями переменных, которые задают это состояние. Система совершает **переход** из одного состояния в другое, если описывающие ее переменные изменяются от значений, задающих одно состояние, на значения, которые определяют другое состояние.

Причина, вызывающая переход из состояния в состояние, называется **событием**.

В качестве иллюстрации этих понятий рассмотрим процесс передачи пакетов в сети передачи данных. Состояние сети будем задавать числом пакетов, находящихся в каждый момент времени в сети. Очевидно, что в этом случае переходы из одного состояния в другое связаны с такими событиями, как поступление пакетов в сеть и завершение передачи пакетов при достижении пункта назначения. Поступление очередного пакета в сеть изменяет состояние (увеличивается значение, описывающее состояние) сети на единицу, а при

выходе пакетов из сети – уменьшает состояние на величину, равную числу пакетов, покинувших сеть в рассматриваемый момент времени. Таким образом, в произвольный момент времени сеть передачи данных может находиться в состояниях 0 (в сети нет пакетов), 1 (в сети один пакет), 2 и т.д.

Понятия «система» и «процесс» тесно связаны и часто рассматриваются как эквивалентные понятия, к которым одинаково применимы термины «состояние» и «переход».

1.3.7. Классификация систем и процессов

Для унификации разрабатываемых моделей и методов проектирования систем все многообразие существующих и возможных систем и процессов следует разбить на отдельные классы, обладающие близкими свойствами и отображаемые определенными моделями, т.е. выполнить их классификацию. Классификация обычно выполняется на основе существенных признаков, в качестве которых будем использовать (рисунок 7.):

- характер функционирования, т.е. характер протекающих в системе процессов (детерминированный или случайный);
- способ изменения значений величин, описывающих состояния системы или процесса (непрерывный или дискретный);
- режим функционирования системы (установившийся или неустановившийся).

1. В зависимости от характера протекающих в системах процессов, системы (процессы) делятся на:

- **детерминированные**, поведение которых может быть предсказано заранее, т.е. параметры системы (прежде всего нагрузочные) представляют собой детерминированные величины;
- **стохастические (случайные, вероятностные)**, в которых процессы развиваются в зависимости от ряда случайных факторов, т.е. параметры системы являются случайными.

2. В зависимости от способа изменения значений величин, описывающих состояния, системы и процессы делятся на два класса:

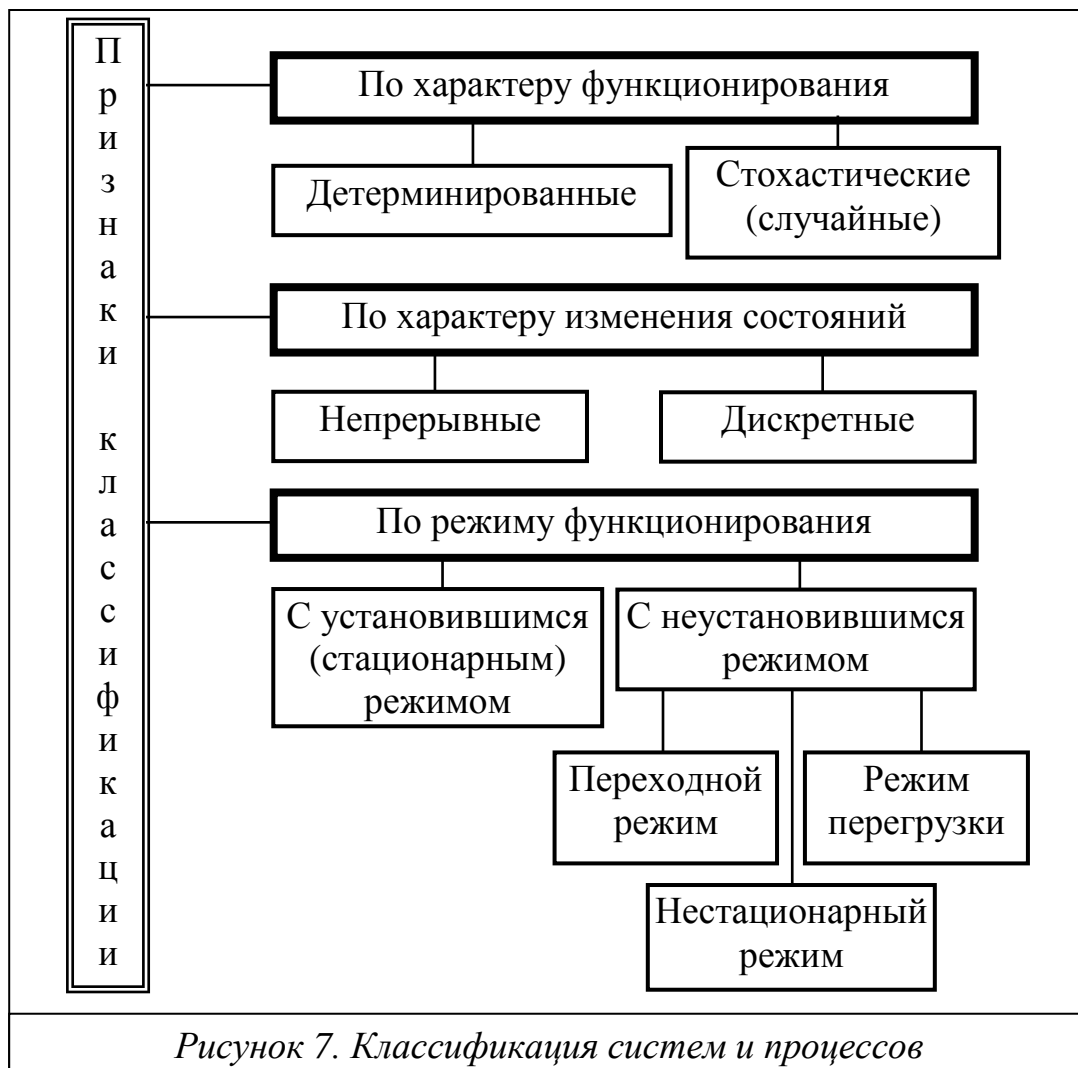
- **непрерывные (с непрерывными состояниями)**, для которых характерен плавный переход из состояния в состояние, обусловленный тем, что величины, описывающие состояние, могут принимать любое значение из некоторого непрерывного интервала (в том числе бесконечного), т.е. являются непрерывными;
- **дискретные (с дискретными состояниями)**, для которых характерен скачкообразный переход из состояния в состояние, обусловленный тем, что величины, описывающие состояния, изменяются скачкообразно и принимают значения, которые могут быть пронумерованы, т.е. являются дискретными, причем число состояний может быть конечным или бесконечным.

3. В зависимости от режима функционирования, системы (процессы) делятся на следующие классы:

- с *установившимся (стационарным) режимом (установившийся или стационарный процесс)*, когда характеристики системы не зависят от времени, т.е. не изменяются со временем;

- с *неустановившимся режимом (процесс неустановившийся)*, когда характеристики системы меняются со временем; неустановившийся режим функционирования системы может быть обусловлен:

- ✓ началом работы системы (переходной режим);
- ✓ нестационарностью параметров системы (нестационарный режим), заключающейся в изменении параметров системы (в первую очередь нагрузочных) со временем;
- ✓ перегрузкой системы (режим перегрузки), когда система не справляется с возложенной на нее нагрузкой.



Большинство исследований технических систем обычно проводится в предположении, что переходной режим завершился, и в системе отсутствуют перегрузки. В противном случае следует использовать специальные методы исследования, которые разрабатываются для переходного режима и режима перегрузок.

Многие реальные системы работают в неустановившемся режиме, обусловленном нестационарностью нагрузки. Для исследования таких систем в

некоторых случаях применимы подходы, позволяющие воспользоваться методами, разработанными для установившегося режима. Во-первых, можно попытаться выделить достаточно продолжительные интервалы времени, в течение которых нагрузка не изменяется, т.е. может считаться стационарной, или же изменение нагрузки незначительно и им можно пренебречь. Во-вторых, исследование, а, следовательно, и проектирование можно проводить методами, разработанными для установившегося режима, в расчете на максимальную или некоторую среднюю нагрузку.

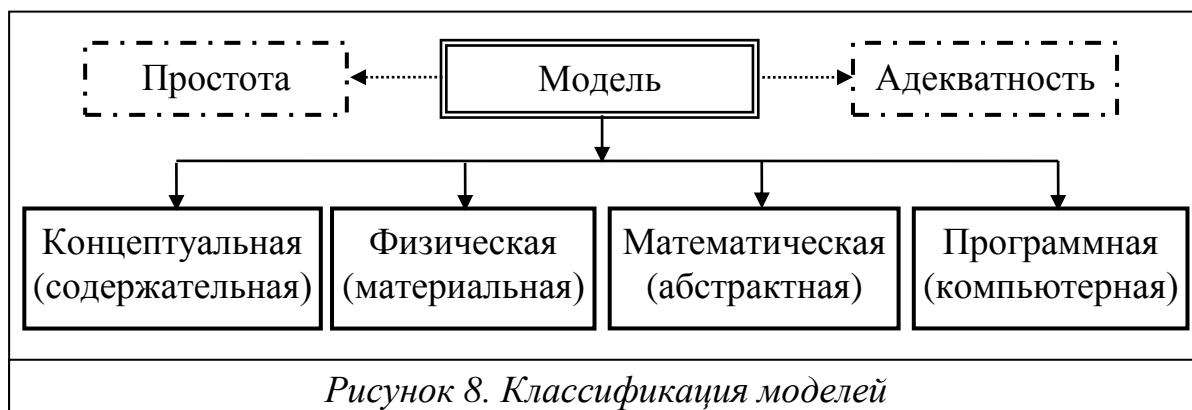
1.4. Модель как средство проектирования

Модель – физический или математический объект, адекватно отображающий исследуемую систему и предназначенный для проведения исследований с целью изучения свойств реальной системы и проектирования новой системы с заданными свойствами.

1.4.1. Типы моделей

В зависимости от способа представления (описания) и реализации модели делятся (рисунок 8) на:

- **концептуальные** или **содержательные**, представляющие собой описание (часто словесное) наиболее существенных особенностей структурно-функциональной организации исследуемой системы, на основе которого разрабатываются физические и математические модели;
- **физические** или **материальные** – модели (макеты), эквивалентные или подобные оригиналу, или модели, процесс функционирования которых такой же, как у оригинала, и имеет ту же или возможно другую физическую природу;
- **математические** или **абстрактные**, представляющие собой формализованное описание системы с помощью математических зависимостей, отражающих процесс функционирования системы в терминах того или иного математического аппарата;
- **программные (алгоритмические, компьютерные)**, представляющие собой программы для ЭВМ, позволяющие наглядно представить исследуемый объект посредством имитации или графического отображения математических зависимостей, описывающих исследуемую систему.



Соответственно различают физическое, математическое и компьютерное моделирование.

При проектировании и исследовании сложных технических систем в настоящее время наиболее широко применяется математическое моделирование, что, в частности, обусловлено появлением высокопроизводительных вычислительных систем и специальных программных средств имитационного моделирования и автоматизированного проектирования.

1.4.2. Основные требования к модели

Ко всем разрабатываемым математическим моделям, используемым при проектировании систем, предъявляются два противоречивых требования (рисунок 8):

- простота модели;
- адекватность исследуемой системе.

Требование простоты модели обусловлено необходимостью построения модели, которая может быть исследована доступными методами и средствами. Построение сложной модели может привести к невозможности получения конечного результата имеющимися средствами в приемлемые сроки и с требуемой точностью, а также к необходимости использования значительных материальных ресурсов и, как следствие, высокой стоимости проектирования.

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, но в то же время и менее адекватна исследуемой системе.

Адекватность (от лат. *adaequatus* – приравненный, равный) означает соответствие модели оригиналу, характеризуемое степенью близости свойств модели и исследуемой системы.

Адекватность математических моделей зависит от:

- степени полноты и достоверности сведений об исследуемой системе, точности представления параметров структурно-функциональной организации и создаваемой в системе нагрузки;

- уровня детализации модели.

При этом моделирование может проводиться:

- в условиях полной определенности, означающей наличие точной информации обо всех исходных структурно-функциональных и нагрузочных параметрах;

- в условиях неопределенности, обусловленных:

- ✓ неточностью сведений о параметрах;
- ✓ отсутствием сведений о значениях некоторых параметров.

Достижение разумного компромисса между простотой модели и ее адекватностью исследуемой системе является одной из сложнейших проблем проектирования. Действительно, с одной стороны, желательно иметь модель с максимальной степенью детализации, отражающую все особенности структурно-функциональной организации исследуемой системы. С другой

стороны, такая модель может оказаться настолько сложной, что ее исследование и использование для целей проектирования будет невозможным или же потребует неоправданно больших материальных и временных ресурсов. Следует также учитывать, что для сложных моделей практически невозможно разработать точные математические методы, а применение громоздких приближенных методов может привести к значительным погрешностям результатов.

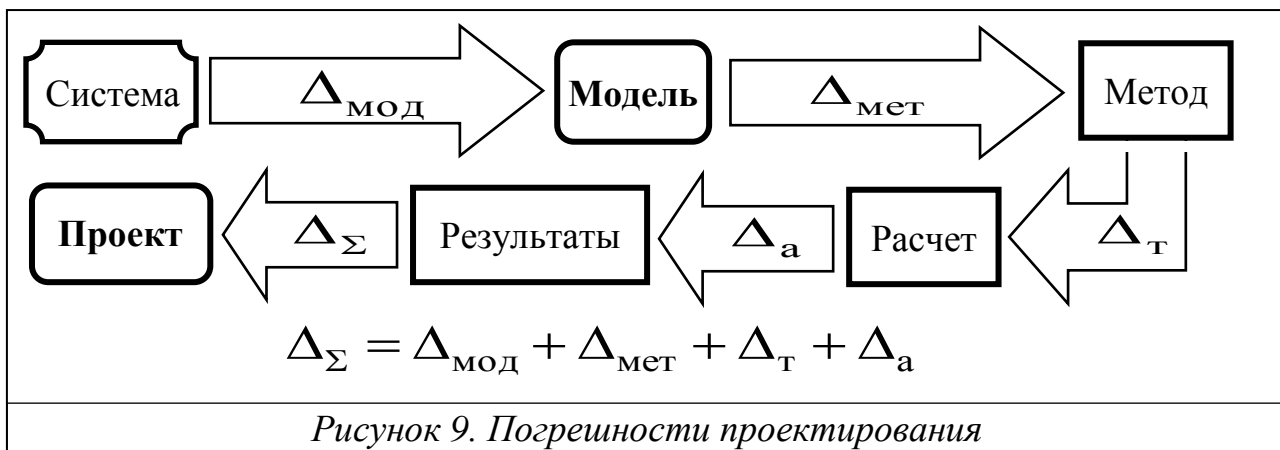
Кроме того, расчет сложных моделей всегда связан с построением громоздких математических зависимостей или же может быть выполнен только средствами имитационного моделирования, что делает чрезвычайно проблемным решение задачи оптимального синтеза, а в некоторых случаях вообще невозможным. В этой ситуации представляется целесообразной разработка более простых моделей и методов расчета, которые, возможно даже с некоторой значительной погрешностью, позволят приближенно решить задачу оптимального синтеза. Впоследствии качество такого решения может быть оценено с использованием адекватной имитационной модели, построенной с высокой степенью детализации. При этом значительно сокращается трудоемкость поиска наилучшего (в идеальном случае – оптимального) решения задачи синтеза за счет существенного уменьшения возможных вариантов структурно-функциональной организации проектируемой системы.

При разработке моделей следует также учитывать точность представления исходных данных, особенно нагрузочных параметров, которые чаще всего неизвестны, либо заданы ориентировочно. Если погрешность представления нагрузочных или структурно-функциональных параметров велика, то, очевидно, нет смысла строить сверхточную модель.

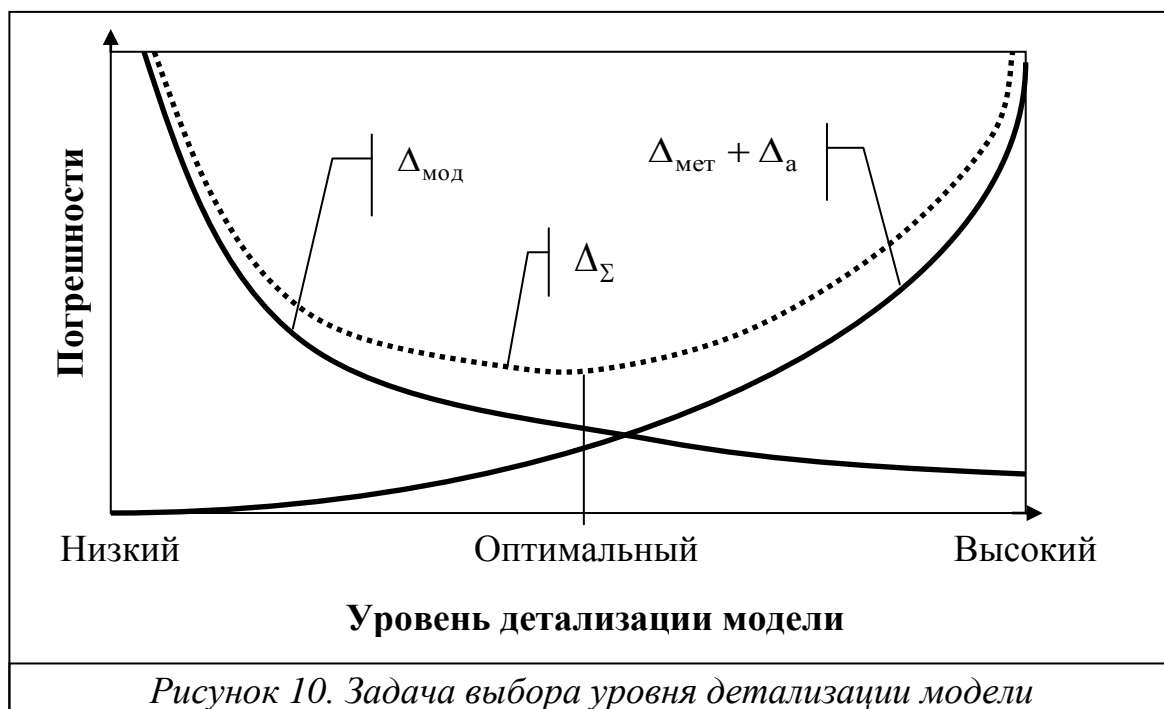
Погрешность результатов моделирования Δ_{Σ} , определяющая качество спроектированной системы, складывается из следующих составляющих (рисунок 9):

- модельная погрешность, зависящая от степени детализации (неадекватности) разработанной модели – $\Delta_{\text{мод}}$;
- методическая погрешность, т.е. погрешность метода, используемого для получения результатов – $\Delta_{\text{мет}}$;
- трансформированная погрешность, обусловленная погрешностью (неточностью) представления исходных данных, преобразуемой (трансформируемой) в процессе расчета в погрешность результатов – $\Delta_{\text{т}}$;
- арифметическая погрешность, связанная с погрешностью округления при проведении расчетов (например, из-за ограниченной длины разрядной сетки компьютера), которая увеличивается с ростом объема вычислений – $\Delta_{\text{а}}$.

Таким образом, суммарная погрешность, вносимая в разрабатываемый проект, равна $\Delta_{\Sigma} = \Delta_{\text{мод}} + \Delta_{\text{мет}} + \Delta_{\text{т}} + \Delta_{\text{а}}$.



С увеличением сложности (адекватности) модели модельная погрешность уменьшается, но при этом растут методическая (из-за применения приближенных методов расчета) и арифметическая (из-за увеличения объема вычислений) погрешности, как это показано на рисунке 10. При этом достаточно сложно оценить, как меняется трансформированная погрешность. Положим, что она изменяется незначительно. Тогда, как видно из представленного графика (рисунок 10), можно говорить о существовании оптимального уровня детализации модели, при котором суммарная погрешность минимальна и достигается компромисс между требованиями адекватности и простоты модели. Определение оптимального уровня детализации модели – сложнейшая задача, от решения которой в значительной степени зависит качество спроектированной системы и которая требует большого опыта и высокой квалификации разработчика.



1.4.3. Классификация математических моделей

Проектирование сложных технических систем с использованием математических моделей, которые разрабатываются на основе концептуальной модели, позволяют получать конкретные значения параметров проектируемой системы, обеспечивающие требуемое качество их функционирования.

Многообразие систем, проявляющееся в многообразии их структурно-функциональной организации, определяет использование множества разных математических моделей в зависимости от особенностей, присущих исследуемой системе. Некоторые из этих особенностей положены в основу классификации математических моделей, представленной на рисунке 11.

1. По назначению различают модели:

- **структурные** (статические), предназначенные для отображения и исследования структурных особенностей системы;
- **функциональные** (динамические), предназначенные для исследования процессов функционирования системы во времени;
- **структурно-функциональные**, предназначенные для исследования структурно-функциональных особенностей системы.



2. В зависимости от характера функционирования исследуемой системы могут использоваться модели:

- **детерминированные**, параметры которых, а, следовательно, и характеристики, представляют собой детерминированные величины;
- **стохастические** или вероятностные, параметры и характеристики которых представляют собой случайные величины.

3. В зависимости от характера протекающих в исследуемой системе процессов могут использоваться модели:

- **непрерывные**, отображающие непрерывные во времени процессы;
- **дискретные**, отображающие дискретные процессы, изменяющие свое состояние скачкообразно в дискретные моменты времени.

4. В зависимости от режима функционирования системы могут использоваться:

- **стационарные** модели, в которых характеристики функционирования, называемые стационарными, инвариантны ко времени;
- **нестационарные** модели, отображающие изменение характеристик функционирования со временем.

Между классами систем и моделей не всегда существует однозначное соответствие, например дискретные системы могут быть представлены непрерывными моделями, а детерминированные системы – вероятностными моделями, и наоборот.

1.4.4. Параметризация моделей

Теоретические исследования сложных систем базируются на использовании моделей, отображающих объект исследования в форме, необходимой и достаточной для получения результатов, составляющих цель исследований.

Модель, как и соответствующая ей система, описывается совокупностью величин, которые могут быть разбиты на параметры и характеристики. Состав параметров и характеристик модели определяется составом параметров и характеристик исследуемой системы и может, в идеальном случае, совпадать с ним. В общем же случае составы параметров и характеристик модели и системы различаются, т.к. для модели они формулируются в терминах математического аппарата, который используется при ее построении, а для системы – в терминах соответствующей прикладной области, к которой принадлежит система. В связи с тем, что, в общем случае, параметры и характеристики системы и модели различаются, их принято называть соответственно **системными** и **модельными**.

Поскольку состав и номенклатура системных и модельных параметров и характеристик, в общем случае, различны, возникает необходимость установления однозначного соответствия между значениями системных и модельных параметров и характеристик, которое выполняется на этапе **параметризации** модели.

Этап параметризации модели в процессе исследования реальной системы оказывает существенное влияние на результаты и, в целом, на качество

проектирования. На этом этапе закладывается фундамент адекватности модели исследуемой системе, поскольку именно в процессе параметризации определяются значения исходных параметров, которые будут использованы в модели и обеспечат достоверность получаемых результатов. Ошибки, заложенные при неудачной параметризации, не смогут быть компенсированы даже применением сверхточной (адекватной) модели и точных методов расчета. Более того, ошибки параметризации могут многократно увеличиться и привести к получению абсолютно неправильных значений исследуемых характеристик и, следовательно, к получению некачественного проекта разрабатываемой системы.

Следует также отметить, что на этапе параметризации устанавливается соответствие не только между значениями системных и модельных параметров и характеристик, но и терминологическое соответствие между заданными в терминах конкретной прикладной области понятиями и элементами исследуемой системы и используемыми в соответствующей математической дисциплине понятиями и элементами математической модели. Например, в вычислительной технике при описании компьютера применяются такие понятия и элементы, как задача, программа, данные, процессор, память и т.д. Положим, что в качестве математической модели компьютера используется случайный процесс, для описания которого в теории случайных процессов применяются такие термины и элементы, как состояние, переход, событие, граф переходов, матрица вероятностей переходов и т.д. Выявление и установление соответствия между указанными понятиями и элементами и является одной из задач этапа параметризации.

Фактически, параметризация – это промежуточный этап установления взаимно-однозначного соответствия между концептуальной и математической моделями.

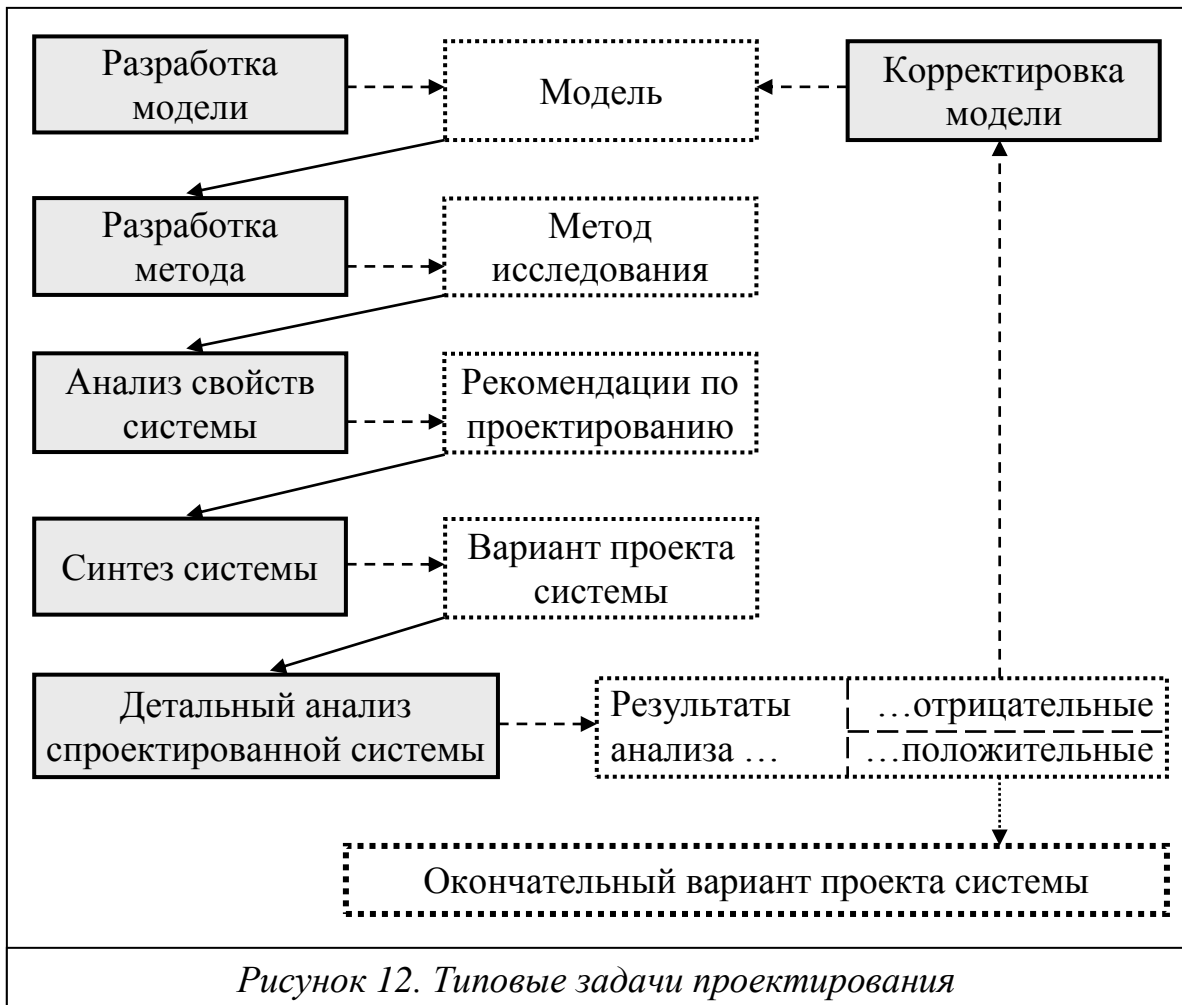
2. Типовые задачи и методы проектирования

2.1. Типовые задачи проектирования

Проектирование любой системы обычно является многоэтапным и на первых этапах предполагает всестороннее исследование её свойств с применением математических моделей. При этом процесс проектирования, в общем случае, является многошаговым и итерационным (рисунок 12), что проявляется в последовательном улучшении проекта системы на каждом шаге.

В процессе выявления и изучения свойств системы, в общем случае, необходимо решать следующие типовые задачи (рисунок 12):

- разработка математической модели;
- разработка метода исследования;
- анализ свойств системы и формирование рекомендаций для проектирования;
- синтез системы и разработка проекта;
- детальный анализ спроектированной системы.



2.1.1. Разработка математической модели

Разработка математической 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.

2.1.2. Разработка метода исследования

Разработка метода исследования, прежде всего, предполагает получение ответа на вопрос: аналитический или имитационный подход будет использоваться для решения задач проектирования – анализа свойств и синтеза системы, удовлетворяющей заданным требованиям к качеству функционирования.

Если в процессе проектирования решается задача выбора варианта построения системы из нескольких возможных, то предпочтение может быть отдано методам имитационного моделирования. Если же проектирование

направлено на разработку наилучшей системы в соответствии с заданным (выбранным) критерием эффективности, т.е. решается задача оптимального синтеза, то более предпочтительными являются методы аналитического моделирования, позволяющие получить математические зависимости характеристик функционирования системы от параметров структурно-функциональной организации и нагрузки. При этом, в зависимости от сложности модели, может быть выбран некоторый существующий метод расчета характеристик функционирования подобных систем, либо может возникнуть необходимость разработки новых аналитических методов исследования.

2.1.3. Анализ свойств системы

Анализ характеристик системы с использованием разработанной модели заключается в выявлении свойств и закономерностей, присущих процессам, протекающим в системах с различной структурно-функциональной организацией, и выработке рекомендаций для решения основной задачи системного проектирования – задачи синтеза. В процессе анализа строятся графические зависимости характеристик функционирования системы от параметров и выявляются наиболее существенные факторы, влияющие на качество функционирования системы.

2.1.4. Синтез системы и разработка проекта

Синтез системы заключается в определении параметров, удовлетворяющих заданным требованиям к качеству функционирования системы, задаваемым в виде ограничений, налагаемых на значения основных характеристик системы. Состав основных характеристик формируется в зависимости от назначения проектируемой системы. Например, если проектируемая система должна обладать высокой надежностью, в качестве основных характеристик может использоваться вероятность безотказной работы, а при проектировании высокоскоростной системы – ее производительность.

Решение задачи синтеза связано с определением зависимостей характеристик функционирования системы от параметров, которые представляются сложными математическими конструкциями. При этом возможность получения приемлемых результатов в процессе решения задач синтеза из-за их сложности и большой трудоемкости, с учетом специфических особенностей реальных систем, превосходит возможности математических методов оптимизации, и задача синтеза в общем виде оказывается математически неразрешимой.

Понятия «синтез» и «проектирование» – достаточно близкие по смыслу и часто используются как синонимы. В то же время между ними существует различие, вытекающее, прежде всего, из их иностранного происхождения.

Термин «синтез», означающий соединение различных элементов в единое целое – систему, неразрывно связан с термином «анализ».

Термин «проектирование» означает процесс создания проекта – прототипа новой системы.

В процессе проектирования технических систем основная задача заключается в создании проекта – комплекта документов, на основе которого строится реальная система, а в процессе синтеза – только определяются параметры и состав проектируемой системы, которые в окончательном проекте могут значительно отличаться от «синтезированных». Таким образом, синтез можно рассматривать как один (может быть даже основной) из этапов проектирования реальных систем.

Можно также считать, что «синтез» – понятие математическое, которое часто используется в таком сочетании как «оптимальный синтез», а «проектирование» – понятие скорее техническое и не всегда предполагает применение каких-то математических методов для построения системы. Другими словами, синтез технических систем реализуется с использованием математических методов моделирования, в то время как проектирование прежде всего, ориентировано на применение различных инженерно-технических решений, обоснование которых может осуществляться математическими расчетами.

2.1.5. Детальный анализ спроектированной системы

Детальный анализ спроектированной системы проводится с целью оценки качества решения задачи проектирования и корректности полученных в процессе синтеза параметров системы, а также выявления предельных возможностей системы, узких мест в системе и т.д.

Поскольку задача синтеза обычно решается на моделях, использующих упрощающие решение предположения и допущения, анализ спроектированной системы, выполняемый с целью определения фактической эффективности спроектированной системы, обычно проводится на основе более детальных моделей, в качестве которых чаще всего используются имитационные или комбинированные аналитико-имитационные модели.

2.1.6. Корректировка модели

Если в процессе детального анализа спроектированной системы будет установлено, что рассматриваемый вариант проекта не удовлетворяет заданным требованиям в отношении характеристик функционирования, необходимо скорректировать модель, которая использовалась для оптимального синтеза и выполнить повторно перечисленные выше задачи. Такая процедура может быть реализована несколько раз до тех пор, пока не будет достигнут требуемый результат.

Корректировка модели и повторное решение задач анализа и синтеза может не потребоваться в тех случаях, когда значения характеристик спроектированной системы незначительно отличаются от заданных значений. Вполне возможно, что для удовлетворения заданных ограничений достаточно внести незначительные изменения в рассматриваемый вариант проекта.

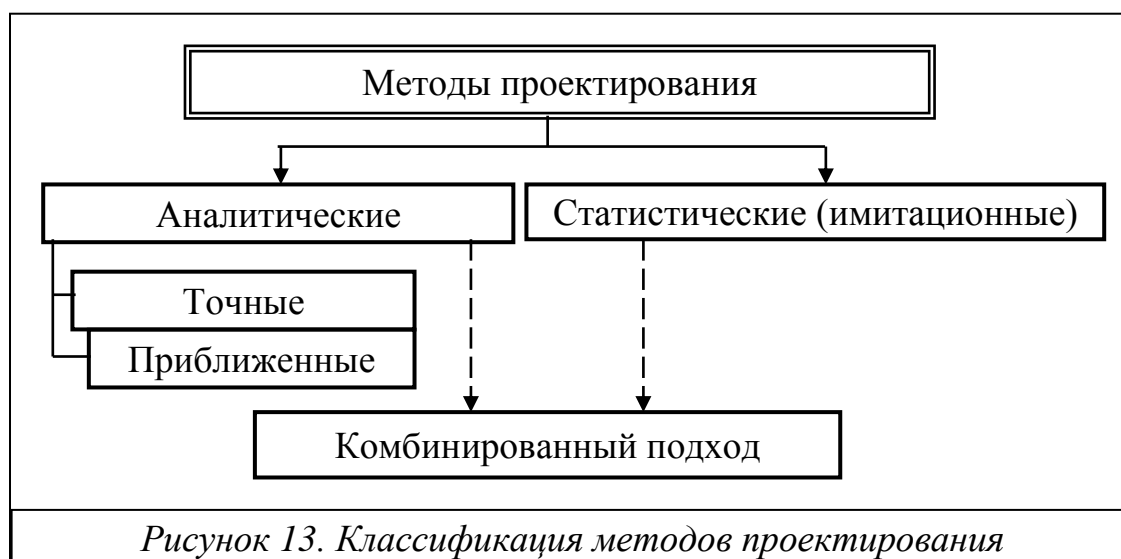
Для того чтобы снизить сложность перечисленных задач, проектирование разделяют на последовательность этапов, на каждом из которых решаются частные задачи проектирования – определяются параметры, связанные с отдельными аспектами организации системы, с использованием тех или иных моделей.

2.2. Методы проектирования

Проектирование и исследование систем обычно проводится на количественном уровне с использованием методов математического моделирования, которые можно разбить на две группы (рисунок 13):

- аналитические (аппарат теории вероятностей, теории массового обслуживания, теории случайных процессов, методы оптимизации, ...);
- статистические или имитационные, основанные на проведении численных экспериментов с применением специальных программных средств и языков моделирования.

На практике проектирование реальных систем обычно реализуется на основе *комбинированного подхода*, предполагающего совместное применение аналитических и имитационных методов.



2.2.1. Аналитические методы

Аналитические методы состоят в построении математических зависимостей, в явном виде отражающих взаимосвязь характеристик функционирования системы от структурно-функциональных и нагрузочных параметров. На их основе строятся графические зависимости, позволяющие наглядно проиллюстрировать влияние параметров на характеристики функционирования системы и выявить наиболее существенные свойства исследуемой системы. Полученные результаты позволяют сформулировать рекомендации для проектирования, направленные на уменьшение количества возможных вариантов построения новой системы.

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 решения задачи выбора оптимального варианта структурно-функциональной организации системы на этапе синтеза.

Недостаток аналитических методов – использование целого ряда допущений и предположений в процессе построения математической модели и невозможность, в некоторых случаях, получить решение в явном виде из-за неразрешимости уравнений в аналитической форме, отсутствия первообразных для подынтегральных функций и т.п. В этих случаях используются численные методы анализа.

Аналитические методы делятся на:

- точные, которые разработаны в основном для простых моделей;
- приближенные, к которым относятся также граничные оценки (верхние и нижние), полученные, например, эвристически или на основе множества имитационных экспериментов.

Таким образом, аналитические методы обычно применяются в следующих случаях:

- для выполнения оценочных расчетов на этапе предварительного анализа и проектирования, не требующих высокой точности получаемых результатов;
- для изучения в широком диапазоне изменения параметров свойств и закономерностей, присущих исследуемой системе; полученные результаты могут служить основой для формирования рекомендаций по проектированию систем;
- для решения задач оптимального синтеза при проектировании новых систем.

2.2.2. Статистические (имитационные) методы

В тех случаях, когда анализ математической модели аналитическими методами может оказаться невозможным или нерезультативным из-за чрезмерной трудоемкости или неустойчивости алгоритмов в отношении погрешностей аппроксимации и округления, строится имитационная модель, в которой процессы, протекающие в системе, описываются как последовательности операций над числами, представляющими значения входов и выходов соответствующих элементов. Имитационные методы реализуются на ЭВМ в виде имитационной модели, объединяющей свойства отдельных элементов в единую систему. Производя вычисления, порождаемые имитационной моделью, можно на основе свойств отдельных элементов определить свойства всей системы, описываемые соответствующими показателями эффективности 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.

Основное достоинство имитационного моделирования заключается в универсальности, т.е. в возможности исследования теоретически систем любой сложности с любой степенью детализации. Применительно к моделированию систем со стохастическим характером функционирования эта универсальность проявляется в возможности исследования свойств систем при любых законах распределения случайных величин, описывающих нагрузку.

Единственным фактором, ограничивающим применение имитационного моделирования, является производительность компьютера, на котором выполняются имитационные эксперименты. Естественно, чем сложнее исследуемая система, чем больше в ней элементов и связей, тем более мощный требуется компьютер, в пределе возможно даже суперЭВМ. При этом мощность компьютера подразумевает не только скорость процессорной обработки, но и большую ёмкость оперативной памяти, а в некоторых случаях – высокие требования к производительности и ёмкости внешней памяти.

В то же время, имитационное моделирование обладает недостатками, ограничивающими его применение. Одним из них является частный характер результатов, не раскрывающий зависимостей характеристик функционирования системы от параметров её структурно-функциональной организации, а лишь определяющий ее в отдельных точках.

Кроме того, имитационное моделирование может служить эффективным инструментом в процессе проектирования только в том случае, если требуется сравнить несколько вариантов построения системы и выбрать из них наилучший. Однако, оказывается практически невозможным (либо это сопряжено с большими временными и материальными затратами) решение задачи оптимального синтеза сложных систем, характеризующихся большой размерностью, то есть наличием большого числа структурно-функциональных и нагрузочных параметров.

Имитационное моделирование обычно используется:

- для установления адекватности аналитических моделей в том случае, если отсутствует возможность сравнения аналитических результатов с результатами измерений на реальной системе;
- для оценки погрешностей приближенных аналитических методов;
- для выбора наилучшего варианта построения (структурно-функциональной организации) системы из нескольких возможных вариантов;
- для детального анализа спроектированной системы.

2.2.3. Комбинированный подход

При исследовании и проектировании сложных систем со стохастическим характером функционирования наиболее универсальным является комбинированный подход, предполагающий совместное применение аналитических и имитационных методов.

С использованием аналитических методов могут решаться задачи проектирования, связанные с формированием требований к структурным и функциональным параметрам, обеспечивающим заданное качество функционирования проектируемой системы, однако получаемые при этом

результаты могут иметь значительную погрешность. Для повышения достоверности приближенных аналитических результатов, полученных в процессе анализа, следует воспользоваться имитационным моделированием, позволяющим оценить погрешности приближенных аналитических зависимостей, а также выявить свойства системы, которые не могли быть получены аналитическими методами, например, свойства системы в случае переходного режима или режима перегрузок.

Аналитические методы позволяют эффективно решать задачу синтеза оптимальной системы. Однако результаты оптимизации из-за применения приближенных аналитических зависимостей могут существенно отличаться от истинных значений. Уточнение результатов оптимизации выполняется на этапе детального анализа спроектированной системы с использованием имитационного моделирования, позволяющего выполнять исследование систем практически любой сложности и с любой степенью детализации.

2.3. Принципы проектирования систем

Проектирование систем на системном уровне базируется на принципах, обеспечивающих корректность и достоверность результатов исследований, полученных с использованием математических моделей, и, в конечном счете, качественное проектирование систем. Такими принципами являются:

- 1) системный подход при решении задач анализа и синтеза;
- 2) принцип иерархического многоуровневого моделирования;
- 3) принцип множественности моделей.

2.3.1. Системный подход и системотехническое проектирование

В основе исследования сложных систем с использованием математического моделирования лежит системный подход, конечной целью которого является системное проектирование, направленное на построение системы с заданным качеством.

Системный подход предполагает, что с учетом свойства интегративности, присущего сложным системам, проектируемая система должна рассматриваться как единое целое. При этом ее проектирование не должно сводиться к независимому проектированию отдельных элементов без учета их взаимосвязи. Это приводит к необходимости применения достаточно больших и сложных моделей, отображающих в целом структурно-функциональную организацию исследуемой системы и протекающие в ней процессы. Соответственно и методы расчета таких моделей являются достаточно сложными и громоздкими и могут быть эффективно реализованы только с применением компьютеров.

Системный подход предполагает решение двух взаимосвязанных задач:

- задачу системного анализа;
- задачу системного проектирования.

В процессе **системного проектирования** необходимо, исходя из сведений о назначении системы и требований, предъявляемых к качеству ее функционирования, определить структурную и/или функциональную

организацию, обеспечивающую реализацию заданных функций. Для этого необходимо располагать знаниями о том, как влияют различные способы структурной и функциональной организации на характеристики функционирования системы, т.е. решать задачи **системного анализа**.

Системное проектирование сложных технических систем, например, вычислительных систем и сетей, называется **системотехническим проектированием**.

В общем виде проблема системотехнического проектирования может формулироваться следующим образом.

Задано назначение системы, определяемое:

- перечнем функций, возлагаемых на систему;
- перечнем и значениями нагрузочных параметров, описывающих взаимодействие системы с внешней средой и потребность в ресурсах системы для реализации заданных функций;
- требованиями к характеристикам системы (мощностным, временным, надежностным, экономическим), которые задаются их предельно допустимыми значениями.

Требуется определить:

- структурную организацию системы, т.е. номенклатуру и состав элементов, а также конфигурацию (топологию) связей между ними;
- функциональную организацию системы, то есть режим функционирования системы, обеспечивающий выполнение заданных ограничений и максимизирующий (минимизирующий) прямой (инверсный) критерий эффективности.

Кроме представленной выше формулировки, системное проектирование может быть направлено на решение частных задач, а именно:

- определение структурной организации системы при заданных параметрах функциональной организации и нагрузки (структурное проектирование);
- определение функциональной организации при заданных параметрах структурной организации и нагрузки (функциональное проектирование);
- определение предельной нагрузки, которая может быть реализована системой с заданной структурно-функциональной организацией (нагрузочное проектирование).

2.3.2. Принцип иерархического многоуровневого моделирования

Качество проектирования системы в значительной мере определяется степенью адекватности математической модели реальной системе, которая достигается за счет использования моделей с разным уровнем детализации, зависящим от особенностей структурно-функциональной организации системы и целей исследования.

Процессы функционирования большинства реальных систем во временном и надежностном аспектах описываются, в общем случае, на основе вероятностного подхода в терминах теории вероятностей с использованием аппарата теории случайных процессов и теории массового обслуживания.

Математические модели позволяют прогнозировать эффект, достигаемый при изменении структурно-функциональных параметров системы и параметров нагрузки.

Процессы функционирования реальных систем практически невозможно описать полно и детально, что обусловлено существенной сложностью таких систем. Основная проблема при разработке модели состоит в нахождении компромисса между простотой ее описания, что является одной из предпосылок понимания и возможности ее исследования аналитическими методами, и необходимостью учета многочисленных особенностей, присущих реальным системам.

Попытка построить единую универсальную модель реальной системы, несомненно, обречена на неудачу ввиду ее необозримости и невозможности расчета. Поэтому проектирование реальных систем целесообразно проводить на основе принципа иерархического многоуровневого моделирования, базирующегося на иерархическом описании исследуемой системы и протекающих в ней процессов.

Принцип иерархического многоуровневого моделирования (ИММ) состоит в следующем. Система и протекающие в ней процессы представляются семейством моделей, каждая из которых описывает поведение системы с точки зрения различных уровней абстрагирования, отличающихся рядом характерных особенностей и параметров, с помощью которых и описывается поведение системы.

Применительно к моделям сложных технических систем можно выделить два направления иерархии (рисунок 14):

- иерархия по вертикали, в которой деление моделей по уровням осуществляется в зависимости от структурно-функциональных особенностей системы;
- иерархия по горизонтали, в которой деление моделей по уровням осуществляется в зависимости от методов их исследования.

В иерархии по вертикали, в общем случае, можно выделить три уровня моделей (рисунок 14):

- уровень **базовых моделей (I)**, представляющих собой простейшие модели, отображающие конкретные специфические особенности структурной или функциональной организации отдельных элементов системы; результаты, полученные на базовых моделях, могут служить основой для разработки и построения более сложных моделей второго и третьего уровней;
- уровень **локальных моделей (II)**, отображающих отдельные аспекты структурно-функциональной организации подсистем или системы в целом и позволяющих решать частные задачи анализа и синтеза;
- уровень **глобальных моделей (III)**, наиболее полно отображающих структурные и функциональные особенности организации системы и представляющих собой модели с высокой степенью детализации.

Глобальные модели строятся на основе результатов, полученных на базовых и локальных моделях.

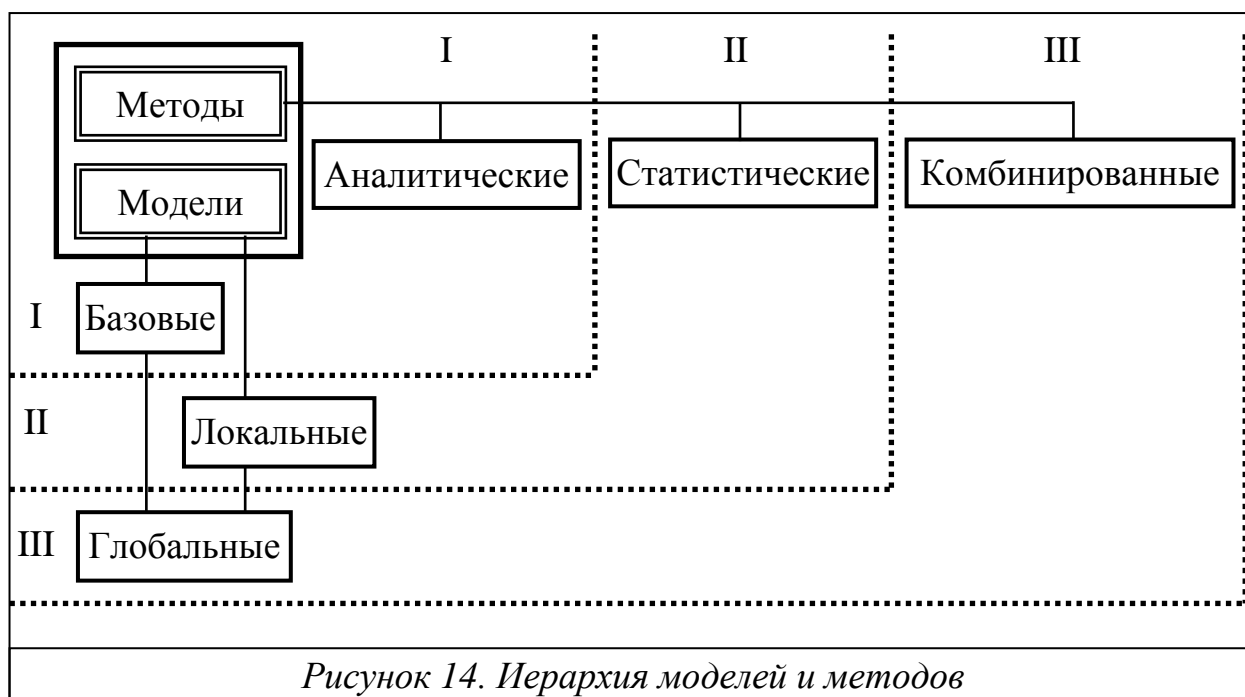


Рисунок 14. Иерархия моделей и методов

Иерархия по горизонтали может включать несколько уровней моделей в зависимости от методов их исследования, например:

- модели, поддающиеся расчету аналитическими методами (уровень I);
- модели, требующие применения статистических методов расчета, основанных на имитационном моделировании (уровень II);
- модели, использующие комбинированные методы расчета (уровень III).

В некоторых случаях уровень моделей, поддающихся расчету аналитическими методами, целесообразно представить в виде двух разных уровней: уровень моделей, поддающихся точному аналитическому расчету, и уровень моделей, поддающихся приближенному аналитическому расчету с приемлемой для инженерных применений точностью, причем результаты могут быть получены либо в явном виде, либо в виде границ (верхней и нижней).

Базовые модели допускают применение точных и приближенных аналитических методов и позволяют получить результат в явном виде. Локальные модели, кроме этого, обычно предполагают применение имитационных методов, а глобальные – наряду с перечисленными методами моделирования, могут использовать комбинированные аналитико-имитационные методы.

Для решения задач проектирования сложных систем обычно применяются модели, поддающиеся точному или приближенному аналитическому расчету. Имитационное моделирование используется для аттестации приближенных методов и детального изучения свойств и закономерностей на моделях большой сложности с целью разработки на основе полученных результатов приближенных и эвристических методов расчета.

Взаимодействие моделей различных уровней иерархии осуществляется путем пересчета характеристик, полученных на одном уровне, в параметры модели, используемой на другом (соседнем) уровне. На каждом уровне может

использоваться множество различных моделей. Состав моделей каждого уровня зависит от структурно-функциональной организации системы и целей исследования. Последнее также определяет степень детализации моделей одного и того же уровня.

Реализация принципа ИММ базируется на методе *структурно-функциональной декомпозиции* проектируемой системы, направленном на выделение и исследование наиболее существенных аспектов структурно-функциональной организации системы.

Структурно-функциональная декомпозиция системы позволяет на разных этапах исследования использовать модели разных уровней:

- на этапе функционального проектирования – базовые модели;
- на этапе структурного проектирования – локальные модели;
- на завершающем этапе структурно-функционального проектирования – глобальные модели.

Такой подход позволяет упростить решение задач системотехнического проектирования, характеризующихся значительной сложностью ввиду большой размерности и громоздкости результатов.

2.3.3. Принцип множественности моделей

Как было сказано выше, математические модели, используемые при проектировании систем, являются абстрактными, что обусловлено переходом от параметров и характеристик реальной системы к её описанию в терминах определённого математического аппарата. В процессе анализа характеристик и исследования свойств математической модели полученные результаты интерпретируются применительно к реальной системе. Абстрактность математической модели состоит в том, что полученные с её помощью результаты могут быть применены к другой реальной системе, которая может быть представлена такой же моделью. Другими словами, одна и та же математическая модель может отображать функционирование совершенно разных по своей природе реальных систем, описываемых с помощью различных структурно-функциональных и нагрузочных параметров, состав и перечень которых диктуется соответствующей прикладной областью. Эта особенность абстрактных моделей лежит в основе принципа множественности моделей, который заключается в следующем. С одной стороны, процессы функционирования различных систем могут быть представлены с помощью одной и той же модели, а, с другой стороны, одна и та же система может быть представлена множеством различных моделей в зависимости от целей исследования. Использование этого принципа позволяет отказаться от подхода, когда для каждой исследуемой системы разрабатывается своя модель, и предложить подход, при котором разрабатываются абстрактные математические модели разного уровня (в основном базовые и локальные), используемые для исследования систем различных классов. При этом одной из важных задач становится задача грамотной параметризации моделей и интерпретации полученных результатов.

2.4. Этапы проектирования систем

Проектирование реальных систем на основе математических моделей в общем случае предполагает выполнение следующих этапов:

- формулировка целей проектирования и требований к разрабатываемой системе;
- разработка концептуальной модели исследуемой системы;
- разработка и параметризация математических моделей системы;
- выбор или разработка методов и средств проектирования;
- проверка адекватности модели (верификация модели);
- проведение экспериментов на модели и анализ характеристик системы;
- решение задачи синтеза;
- детальный анализ спроектированной системы.

2.4.1. Формулировка целей проектирования и требований к разрабатываемой системе

На этом этапе решаются следующие задачи:

- определяются возможные способы структурно-функциональной организации и условия функционирования проектируемой системы;
- формулируются задачи анализа и/или синтеза, которые должны быть решены в процессе проектирования;
- конкретизируются наиболее важные характеристики, подлежащие исследованию, и формы представления результатов проектирования;
- определяется состав показателей эффективности и формулируются требования к качеству функционирования в виде ограничений, налагаемых на характеристики системы;
- определяются требования к точности получения результатов и форма их представления.

Этап формулировки целей проектирования и требований к разрабатываемой системе существенно влияет на все последующие этапы.

2.4.2. Разработка концептуальной модели проектируемой системы

В процессе построения концептуальной модели выявляются причинно-следственные связи, присущие исследуемой системе и существенные для достижения целей проектирования.

Основное назначение концептуальной модели – выявление наиболее существенных аспектов структурно-функциональной организации системы, учет которых необходим для получения требуемых результатов. В концептуальной модели обычно в словесной форме приводятся сведения о природе и параметрах элементарных явлений проектируемой системы, о виде и степени взаимодействия между ними, о месте и значении каждого элементарного явления в общем процессе функционирования системы.

Одна и та же система может представляться различными концептуальными моделями, которые строятся в зависимости от целей

исследования, сформулированных на предыдущем этапе. Например, одна концептуальная модель может отображать временные аспекты функционирования системы, другая – надежность аспекты построения системы.

Построение концептуальной модели и ее формализация предполагает выполнение следующих этапов:

1) постановка задачи моделирования, включающая:

- формулировку целей и обоснование необходимости моделирования;
- определение номенклатуры показателей эффективности в зависимости от целей моделирования;
- оценку размерности задачи и определение возможности ее разбиения на подзадачи;
- формирование требований к составу исходных параметров для проведения моделирования;

• выдвижение гипотез и принятие предположений и допущений;

2) определение исходных параметров и их описание, включающее:

- выбор наименований и обозначений параметров;
- выбор единиц измерения параметров;
- установка диапазонов изменения параметров;

3) формирование критерия эффективности, предполагающее построение обобщенного показателя эффективности на основе множества частных, зачастую противоречивых, показателей с использованием одного из следующих подходов:

• построение **составного критерия эффективности** в виде **аддитивного** F_1 или **мультипликативного** F_2 функционала:

$$F_1 = \sum_{i=1}^N \alpha_i x_i; \quad F_2 = \frac{\prod_{i=1}^n x_i}{\prod_{i=n+1}^N x_i},$$

где x_1, \dots, x_N – частные показатели эффективности; α_i – весовой коэффициент показателя x_i ;

• выбор в качестве критерия эффективности F одного **частного показателя при ограничениях**, налагаемых на остальные показатели эффективности:

$F = x_k$ при ограничениях $x_i \leq x_i^*$ или $x_i \geq x_i^*$ для всех $i \neq k$;

4) описание концептуальной модели системы, предполагающее:

• словесное описание и выявление особенностей структурно-функциональной организации проектируемой системы в терминах и понятиях реальной системы;

• описание модели с использованием типовых математических схем, временных диаграмм или в виде блок-схем алгоритмов;

• окончательное принятие гипотез, предположений и допущений;

5) проверка достоверности концептуальной модели, включающая:

- проверку логики построения и описания модели;
- оценку достоверности исходных данных;
- анализ принятых предположений и допущений.

2.4.3. Разработка и параметризация математических моделей

Концептуальная модель служит основой для разработки математических моделей в терминах конкретного математического аппарата.

Создание математической модели преследует две основные цели:

- дать формализованное описание структуры и процесса функционирования системы для однозначности их понимания;
- попытаться представить процесс функционирования системы в виде, допускающем аналитическое исследование системы с использованием методов и приемов, разработанных в рамках данного математического аппарата.

Выбор того или иного математического аппарата обусловлен физической природой исследуемой системы и процессов, протекающих в ней. Для исследования систем со стохастическим характером функционирования широко применяется аппарат теории случайных процессов и теории массового обслуживания.

Основная проблема при создании модели заключается в нахождении компромисса между простотой модели, что является одной из предпосылок понимания и возможности ее эффективного исследования, и ее адекватностью исследуемой системе.

В связи с тем, что состав и номенклатура системных и модельных параметров и характеристик, в общем случае, различается, возникает необходимость установления соответствия между значениями системных и модельных параметров и характеристик, которое выполняется в процессе параметризации модели.

2.4.4. Выбор или разработка методов и средств проектирования

Проектирование систем предполагает применение следующих математических методов моделирования:

- аналитических;
- статистических (имитационных);
- комбинированных.

Выбор конкретного метода зависит от многих факторов, в том числе от:

- целей проектирования;
- сложности проектируемой системы;
- сложности модели, определяемой выбранным уровнем ее детализации;
- требований к номенклатуре исследуемых характеристик;
- требований к точности получаемых результатов;
- требований к общности получаемых результатов;
- требований к затратам времени на моделирование и проектирование;
- требований к материальным затратам;

- наличия специальных технических средств для проведения проектирования;

- квалификации специалиста, проводящего проектирование и т.д.

Результаты сравнительного анализа методов проектирования, выполненного на качественном уровне, представлены в таблице 1 (фигурными скобками отмечены наилучшие значения каждого показателя), где «в», «с» и «н» - означают высокий, средний и низкий уровень показателя.

Достоинства аналитических методов проектирования, обуславливающие их широкое использование в задачах анализа и синтеза сложных систем:

- возможность проведения всестороннего исследования свойств системы в широком диапазоне изменения параметров;

- общность аналитических методов расчета и получаемых на их основе результатов, проявляющаяся в их применимости для широкого класса систем, а не для одной конкретной;

- возможность решения задач синтеза в условиях большой номенклатуры оптимизируемых параметров (число параметров может достигать нескольких сотен и тысяч) и выбранного критерия эффективности;

- малые затраты времени на получение результатов, что особенно важно на этапе предварительного проектирования, в процессе которого требуется получать приближенные оценки характеристик системы в сжатые сроки;

- возможность решения задач большой (практически неограниченной) размерности, обусловленной большим числом параметров, описывающих структурно-функциональную организацию системы и нагрузку;

- возможность в приемлемые сроки проиграть несколько различных вариантов организации проектируемой системы с целью их сравнительного анализа.

Таблица 1

Метод моделирования	Сложность метода	Общность рез-тов	Точность рез-тов	Затраты времени	Матер. затраты	Задачи синтеза
Аналитический	{н}	{в}	н	{н}	{н}	{в}
Имитационный	с	н	{в}	в	в	н
Комбинированный	в	с	с	с	с	с

Основным достоинством имитационных методов является возможность построения моделей систем любой сложности с любой степенью детализации и, как следствие этого, возможность получения результатов исследования практически с любой наперед заданной точностью.

На практике обычно используется подход, при котором разрабатываются одновременно одна или несколько аналитических и имитационных моделей, при этом имитационные модели применяются как для оценки погрешностей приближенных аналитических моделей, так и для детального анализа синтезированной на основе приближенных аналитических моделей оптимальной системы.

В качестве технических средств проектирования обычно используются средства вычислительной техники: от персональных компьютеров для относительно простых систем до суперЭВМ при проектировании сложных систем.

В качестве программных средств могут быть использованы процедурно-ориентированные или проблемно-ориентированные алгоритмические языки, а также специализированные программные средства автоматизированного проектирования сложных систем, которые можно разбить на три группы:

- коммерческие средства, ориентированные на проектирование конкретных систем в определённой прикладной области и характеризующиеся удобным графическим интерфейсом и наличием библиотеки реальных устройств и оборудования, входящего в состав системы; недостатком коммерческих средств является их закрытость, что не позволяет, при необходимости, расширить диапазон исследований и выполнить более детальный анализ характеристик функционирования системы, а также высокая стоимость, достигающая в некоторых случаях десятков и даже сотен тысяч долларов США; примером коммерческой системы может служить Net Cracker Professional для проектирования компьютерных сетей;

- научно-исследовательские системы и языки моделирования систем различных классов, например, система (язык) имитационного моделирования GPSS;

- специализированные средства в виде программных комплексов, ориентированных на системы (модели) определённого класса (МВС, ИМСС,...).

Технические и программные средства проектирования выбираются с учетом ряда факторов, основными среди которых являются достаточность и полнота средств для реализации концептуальной и математической модели. Среди других факторов можно назвать доступность средств, простота и легкость освоения технических и программных средств проектирования, скорость и корректность создания программной модели, существование методики использования средств для проектирования систем определённого класса.

После выбора средств проектирования разрабатывается программная модель. Этот процесс включает разработку алгоритма, конкретизацию форм представления исходных данных и результатов, написание и отладку программы.

2.4.5. Проверка адекватности модели

Проверка адекватности математической модели исследуемой системе (верификация модели) заключается в анализе ее соответствия исследуемой системе, проявляющегося в близости значений модельных и системных характеристик.

Отличие модели от исследуемой системы связано с тем, что обычно модель является упрощенным и идеализированным отображением исследуемой системы, которое обусловлено:

- идеализацией внешних условий и режимов функционирования;
- не учетом в модели несущественных, по мнению исследователя, факторов и параметров;
- отсутствием точных сведений о внешних воздействиях и о некоторых конкретных нюансах организации системы;
- введением ряда упрощающих предположений и допущений.

Мерой адекватности математической модели исследуемой системе может служить абсолютное Δ или относительное δ отклонение модельных характеристик H_M от системных H_C :

$$\Delta = |H_M - H_C|; \quad \delta = |H_M - H_C| / H_C = \Delta / H_C.$$

Тогда показателем адекватности может служить вероятность того, что отклонение Δ не превышает некоторого предельного значения Δ^* : $\pi = \Pr(\Delta < \Delta^*)$.

Однако применение данного показателя на практике затруднено, а во многих случаях и невозможно по следующим причинам:

- для проектируемых или модернизируемых систем обычно заранее неизвестны значения системных характеристик H_C ;
- система в большинстве случаев оценивается по множеству системных характеристик, которые могут иметь разные значения отклонений Δ и разную размерность.

На практике верификация модели обычно проводится путем экспертного анализа разумности результатов моделирования.

Можно выделить следующие этапы верификации модели:

- проверка элементов модели и правильности формирования значений их параметров, особенно задаваемых в виде случайных величин;
- проверка адекватности формирования нагрузки в модели;
- проверка концептуальной модели с целью выявления ошибок постановки задачи;
- проверка математической модели с целью выявления ошибок математического описания структурно-функциональной организации системы и нагрузки;
- оценка точности приближенных аналитических методов расчета характеристик модели;
- проверка программной модели с целью выявления логических ошибок в алгоритме и инструментальных ошибок в программе.

В случае выявления неадекватности модели исследуемой системе необходимо выполнить корректировку или калибровку модели, которая может быть: глобальной и локальной.

Глобальная корректировка заключается в разработке новой модели и необходима при наличии:

- ошибок в постановке задачи моделирования;

- методических ошибок в концептуальной или математической модели системы.

Локальная корректировка может состоять:

- в уточнении параметров модели;
- в изменении метода расчета характеристик;
- в разработке более детализированной математической модели;
- в изменении программной модели.

В процессе проверки адекватности модели необходимо определить область применения модели, т.е. оценить диапазон изменения параметров, при котором точность результатов моделирования находится в допустимых пределах.

2.4.6. Проведение экспериментов на модели и анализ характеристик системы

Исследования на моделях заключаются в проведении экспериментов, в процессе которых определяются характеристики системы при разных значениях структурно-функциональных параметров и параметров нагрузки. Большая номенклатура исходных параметров и широкий диапазон их изменения требует предварительного планирования выполняемых на модели экспериментов (расчетов).

Планирование направлено на уменьшение длительности эксперимента при условии обеспечения достоверности и полноты результатов моделирования.

Особую значимость планирование экспериментов приобретает при использовании методов имитационного моделирования, характеризующихся большими затратами ресурсов ЭВМ в процессе моделирования. Одной из основных проблем имитационного моделирования является нахождение компромисса между временем моделирования и затратами памяти ЭВМ, на которой проводится моделирование, связанного с тем, что имитационное моделирование предъявляет повышенные требования как к производительности, так и к памяти ВМ для проведения имитационных экспериментов. Время, затрачиваемое на проведение одного эксперимента с моделью средней сложности даже на высокопроизводительных ВС может достигать нескольких десятков минут и, в некоторых случаях, нескольких часов, а потребность в оперативной памяти ВС – до нескольких десятков и сотен гигабайт. Причем с увеличением числа проводимых имитационных экспериментов соответственно возрастает время моделирования. Все это обуславливает высокую стоимость имитационного моделирования и требует тщательного планирования имитационных экспериментов с целью сокращения затрат на моделирование.

При проведении имитационного моделирования используются два способа планирования:

- **стратегическое планирование**, состоящее в выборе определенных сочетаний параметров и последовательности проведения экспериментов с использованием методов теории планирования экспериментов;

• **тактическое планирование**, направленное на уменьшение времени выполнения одного эксперимента при обеспечении статистической достоверности результатов имитационного моделирования.

Анализ результатов моделирования направлен на выявление свойств, присущих исследуемой системе, и включает в себя следующие этапы:

- обработка результатов для удобства последующего анализа и использования; на этом этапе выделяются наиболее важные с точки зрения исследователя и с учетом целей исследования результаты, которые представляются в форме, наиболее удобной для изучения свойств исследуемой системы;

- определение зависимостей характеристик от параметров системы путем варьирования исходных параметров структурно-функциональной организации и нагрузки с целью выявления и формулирования свойств исследуемой системы;

- принятие решения о работоспособности исследуемой системы и выработка рекомендаций по наиболее эффективной и рациональной организации проектируемой или модернизируемой системы, которые могут быть использованы в дальнейшем при решении задач синтеза в процессе системотехнического проектирования.

2.4.7. Решение задачи синтеза

Синтез оптимальной системы направлен на построение системы, наилучшим образом соответствующей своему назначению.

Основные этапы синтеза:

- формирование критерия эффективности, устанавливающего способ оценки качества системы в целом;

- определение оптимальных параметров структуры и режима функционирования системы, обеспечивающих выполнение заданных ограничений на характеристики системы (задача оптимизации);

- анализ качества функционирования оптимальной системы путем сопоставления её характеристик с заданными ограничениями на имитационных моделях или реальных системах.

Решение задачи синтеза связано с определением зависимостей характеристик функционирования системы от параметров, которые представляются сложными математическими конструкциями. При этом возможность получения приемлемых результатов в процессе решения задач синтеза из-за их сложности и большой трудоемкости с учетом специфических особенностей реальных систем превосходит возможности математических методов оптимизации, и задача синтеза в общем виде оказывается математически неразрешимой. Для того чтобы снизить сложность задачи синтеза, процесс проектирования разделяют на последовательность этапов, на каждом из которых решаются частные задачи синтеза – определяются параметры, связанные с отдельными аспектами организации системы, с использованием тех или иных моделей.

В зависимости от целей можно выделить следующие частные задачи (этапы) синтеза:

• **структурный синтез**, состоящий в выборе способа структурной организации системы, в рамках которой могут быть удовлетворены требования технического задания; структурный синтез включает в себя два этапа:

- ✓ элементный синтез, состоящий в определении требований к параметрам отдельных элементов системы;
- ✓ топологический (конфигурационный) синтез, состоящий в определении способа взаимосвязи элементов системы, т.е. топологии (конфигурации) системы;

• **функциональный синтез**, состоящий в выборе режима (способа) функционирования системы;

• **нагрузочный синтез**, состоящий в определении требований к параметрам нагрузки, обеспечивающим функционирование системы с заданным качеством.

На каждом из перечисленных этапов синтеза определяются значения соответствующего подмножества параметров, характеризующих структурную, функциональную организацию системы или нагрузку, возлагаемую на систему. При этом значения параметров оптимизируются лишь в отношении факторов, учитываемых на каждом из этапов синтеза, но не в отношении системы в целом. Поэтому многоэтапный синтез позволяет получить лишь приближенные оптимальные решения, качество которых проверяется путем детального анализа синтезированной системы.

2.4.8. Детальный анализ спроектированной системы

Необходимость детального анализа спроектированной системы обусловлена необходимостью получения качественного проекта синтезируемой системы. Задача синтеза обычно решается с использованием сравнительно простых моделей, позволяющих получить решение в явной аналитической форме. При этом погрешность модели, а также методов расчета характеристик системы в случае применения приближенных аналитических зависимостей может привести к значительным различиям между расчетными и реальными значениями оптимизируемых параметров. В связи с этим возникает необходимость проверки и уточнения найденных значений параметров структурно-функциональной организации системы, для чего необходимо использовать наиболее адекватные модели, позволяющие получить результаты, в максимальной степени соответствующие реальным. В качестве таких моделей обычно применяются имитационные модели, которые могут быть построены с максимальным приближением к реальной системе за счёт большей детализации по сравнению с аналитической моделью.

Кроме того, в процессе детального анализа синтезированной системы должны быть выявлены предельные возможности системы, узкие места в системе, а также определено, насколько хорошо (с каким запасом) выполняются заданные требования к качеству функционирования проектируемой системы.

3. Математические модели дискретных систем

Многообразие структурно-функциональной организации реальных систем определяет использование множества разных моделей для решения задач системного проектирования.

Случайный характер процессов, протекающих в подавляющем большинстве реальных систем (например, процессы обработки и передачи данных в компьютерных сетях) обуславливает широкое применение *дискретных стохастических (вероятностных) моделей*, функционирование которых описывается случайными величинами.

Для рассматриваемых ниже моделей будем полагать, что:

- априорно известны параметры, описывающие структурную и функциональную организацию системы и создаваемую в ней нагрузку;
- система и, следовательно, модель функционирует в установившемся режиме, при котором её характеристики не меняются со временем;
- модели относятся к классу структурно-функциональных моделей, отображающих структурные и функциональные особенности организации проектируемой системы.

В зависимости от способа представления и реализации будем использовать концептуальные (содержательные), математические (абстрактные) и программные (компьютерные) модели.

Для изучения процессов, протекающих в дискретных системах со стохастическим характером функционирования, широко используются **модели массового обслуживания (ММО)**, которые делятся на **простейшие (базовые) модели** в виде систем массового обслуживания и **сетевые модели** в виде сетей массового обслуживания, представляющие собой математические объекты, описываемые в терминах математического аппарата теории массового обслуживания.

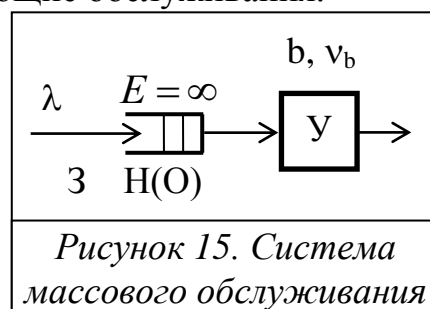
3.1. Базовые модели

В качестве простейших математических моделей систем со стохастическим характером функционирования используются **системы массового обслуживания (СМО)**, содержащие один (рисунок 15) или несколько (рисунок 17) устройств $У$ (устройств, приборов, каналов), обслуживающих заявки (запросы) $З$, поступающие в систему, и один (рисунок 15 и 17) или несколько (рисунок 17) накопителей $Н$, в которых находятся заявки, образующие очереди ($О$) и ожидающие обслуживания.

СМО, используемые в качестве моделей дискретных систем со стохастическим характером функционирования, могут быть классифицированы по следующим признакам:

1) по числу мест в накопителе:

- с накопителем неограниченной емкости $E = \infty$ (рисунок 15), называемые СМО без потерь;



- с накопителем ограниченной емкости $E < \infty$ (рисунок 17), называемые СМО с потерями;

- без накопителя ($E = 0$), называемые СМО с отказами;

2) по количеству устройств:

- с одним устройством (рисунок 15), называемые одноканальными;

- с несколькими устройствами (рисунок 17), называемые многоканальными, причем, устройства могут быть идентичными (однородные системы) и затрачивать одинаковое время на обслуживание любой заявки, либо разными (неоднородные системы), длительности обслуживания заявок в которых различны; в дальнейшем, если не оговорено другое, будем полагать, что все устройства идентичны и равнодоступны для любой заявки;

3) по количеству классов (типов) заявок, поступающих в СМО:

- с однородным потоком заявок одного класса (рисунок 15 и 17);

- с неоднородным потоком заявок нескольких классов (рисунок 17).

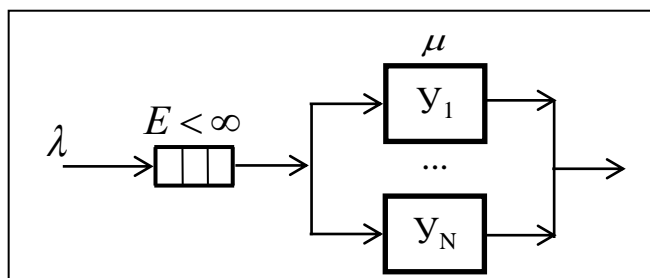


Рисунок 17. СМО с N устройствами и накопителем ограниченной емкости

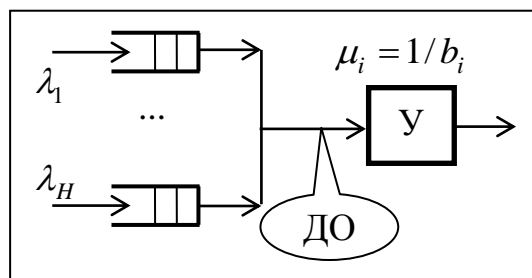


Рисунок 17. СМО с неоднородным потоком

Следует отметить, что заявки в математической модели будут отнесены к разным классам только в том случае, если они различаются *длительностью обслуживания* в устройстве и/или имеют *разные приоритеты* при занесении в накопитель или при выборе из накопителя на обслуживание. В остальных случаях заявки образуют однородный поток.

Модели реальных систем, представляемые в виде СМО разных классов, будем называть **базовыми моделями**.

Для описания базовых моделей используются три группы параметров:

- структурные;
- нагрузочные;
- функциональные параметры (параметры управления).

К структурным параметрам относятся:

- количество устройств N ;
- количество H и ёмкости E_1, \dots, E_H накопителей;

- способ взаимосвязи накопителей с устройствами, задаваемый, например, в виде матрицы инцидентностей (связей) $I = \{i_{h,n} \mid h = \overline{1, H}; n = \overline{1, N}\}$, элементы которой принимают два значения: $i_{h,n} = 1$, если накопитель h связан с устройством n , и $i_{h,n} = 0$ – в противном случае.

Нагрузочные параметры СМО включают в себя параметры потока заявок, поступающих в систему, и параметры обслуживания заявок. К ним же относится количество классов заявок N , которое равно 1 для СМО с однородным потоком заявок и $N > 1$ для СМО с неоднородным потоком.

Функциональные параметры задаются в виде конкретных стратегий управления потоками заявок в СМО, определяющих правила занесения заявок в накопители и выбора из очереди на обслуживание.

3.1.1. Параметры потока заявок

Основным параметром потока заявок является его *интенсивность* λ – среднее число заявок, поступающих в систему за единицу времени. Величина $a = 1/\lambda$ определяет средний интервал времени между двумя последовательными заявками.

Для описания детерминированного (регулярного) потока заявок достаточно задать интенсивность потока λ или значение интервала $a = 1/\lambda$.

Для случайного потока заявок, в общем случае, необходимо задать законы распределений $A(\tau)$ интервалов времени τ между заявками.

При исследовании моделей массового обслуживания аналитическими методами часто предполагается, что поток заявок – простейший. Предположение о простейшем потоке заявок позволяет для многих математических моделей получить достаточно *простые* аналитические зависимости характеристик от параметров *в явном виде*. Свое название «простейший» поток получил именно благодаря этой особенности. Анализ моделей с потоками заявок, отличными от простейших, обычно усложняет математические выкладки и в большинстве случаев не позволяет получить аналитическое решение в явном виде. В случае потоков, отличных от простейшего, эффективным инструментом исследования становится имитационное моделирование.

Простейший поток – это стационарный ординарный поток без последствия.

Простейший поток иногда называют «экспоненциальным», поскольку интервалы времени τ между заявками в таком потоке распределены по экспоненциальному закону с функцией и плотностью распределения соответственно

$$A(\tau) = 1 - e^{-\lambda\tau}, \quad a(\tau) = \lambda e^{-\lambda\tau}, \quad (1)$$

где $\lambda > 0$ – интенсивность потока заявок.

Легко убедиться, что среднее значение (математическое ожидание) и дисперсия интервалов между заявками в простейшем потоке соответственно равны $a = 1/\lambda$ и $D = 1/\lambda^2$, тогда коэффициент вариации $\nu = \sqrt{D}/a = 1$.

Простейший поток часто называют **пуассоновским**, поскольку число заявок, поступающих за некоторый заданный промежуток времени t , распределено по закону Пуассона :

$$P(k, t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad (2)$$

где $P(k, t)$ – вероятность поступления k заявок за некоторый фиксированный интервал времени t ($k = 0, 1, \dots; t > 0$); λ – интенсивность потока заявок.

Простейший поток заявок обладает следующими замечательными свойствами.

1. Сумма H простейших потоков с интенсивностями $\lambda_1, \dots, \lambda_H$ образует простейший поток с интенсивностью

$$\Lambda = \sum_{k=1}^H \lambda_k. \quad (3)$$

2. **Вероятностное** (но не детерминированное) разрежение простейшего потока заявок, при котором любая заявка случайным образом с некоторой вероятностью p исключается из потока независимо от того, исключены другие заявки или нет, приводит к образованию простейшего потока с интенсивностью $\lambda' = p\lambda$, где λ – интенсивность исходного потока. Поток исключенных заявок – также простейший с интенсивностью $\lambda'' = (1-p)\lambda$.

Простейший поток заявок является математическим представлением некоторого «идеального» потока, обладающего рядом замечательных свойств, благодаря которым для многих математических моделей удаётся получить достаточно простые аналитические зависимости, связывающие характеристики функционирования систем массового обслуживания с исходными параметрами. Одним из таких свойств является «отсутствие последствия», которое заключается в том, что поступление в систему очередной заявки не зависит от того, когда и сколько заявок поступило ранее.

Предположение о простейшем потоке широко используется не только из-за простоты получения математических зависимостей, но и по той причине, что многие реальные потоки близки к простейшим. Эта близость во многих случаях обусловлена следующим.

Во-первых, суммирование (объединение) независимых стационарных ординарных потоков образует простейший поток при условии, что складываемые потоки оказывают более или менее одинаковое влияние на суммарный поток, причем на практике суммарный поток становится близким к простейшему уже при суммировании 5 потоков. Отметим, что при этом к суммируемым потокам не предъявляется требование отсутствия последствия.

Во-вторых, можно показать, что стационарный ординарный поток заявок близок к простейшему, если на него оказывает влияние множество случайных факторов.

Предположение о простейшем характере входного потока заявок оправдано также в тех случаях, когда известно, что коэффициент вариации интервалов между последовательными заявками реального потока меньше единицы. В этом случае использование простейшего потока в модели позволяет получить так называемые верхние оценки характеристик обслуживания заявок,

гарантирующие, что в реальной системе значения характеристик будут не хуже, чем полученные на модели.

3.1.2. Параметры обслуживания заявок

Длительность обслуживания заявок в устройстве в общем случае задается функцией $B(\tau)$ или плотностью $b(\tau) = B'(\tau)$ распределения.

Величина, обратная средней длительности обслуживания b , характеризует среднее число заявок, которое может быть обслужено за единицу времени, и называется **интенсивностью обслуживания**: $\mu = 1/b$.

Во многих случаях аналитические зависимости могут быть получены для произвольного закона распределения длительности обслуживания заявок, если поток поступающих в систему заявок – простейший. При этом для определения средних значений характеристик обслуживания достаточно задать, кроме математического ожидания b , второй момент распределения или коэффициент вариации ν_b длительности обслуживания.

Время T_0 , измеряемое от момента поступления некоторой заявки в систему до момента завершения обслуживания заявки, находящейся в устройстве, называется **временем дообслуживания**. Математическое ожидание этого времени

$$M[T_0] = \rho b(1 + \nu_b^2)/2, \quad (4)$$

где λ – интенсивность простейшего потока заявок, поступающих в систему.

Время T_0 учитывает, что в момент поступления заявки в системе может и не оказаться заявок, т.е. учитывает простой системы.

3.1.3. Стратегии управления потоками заявок

В современных системах с очередями, например в компьютерных сетях, широко применяются разнообразные стратегии управления потоками заявок и организации очередей, определяющих:

- правило занесения поступающих заявок в накопитель (буфер), называемое **дисциплиной буферизации**;
- правило выбора заявок из очереди для обслуживания в устройстве, называемое **дисциплиной обслуживания**.

Реализация этих правил осуществляется с учетом **приоритетов** заявок, устанавливающих преимущественное право заявок одного класса по отношению к заявкам других классов на занесение (в накопитель) или выбор из очереди (для обслуживания в устройстве).

Дисциплины буферизации могут быть разбиты на два класса:

- без вытеснения заявок, находящихся в накопителе, при этом заявки, поступившие в систему и заставшие накопитель заполненным до конца, теряются;
- с вытеснением заявки данного класса или заявки самого низкоприоритетного класса, которая теряется.

В последнем случае дисциплины буферизации могут использовать следующие правила вытеснения заявок такого же или самого низкого приоритета из накопителя:

- вытеснение последней заявки, т. е. поступившей в систему позже всех;
- вытеснение «долгой» заявки, т. е. находящейся в накопителе дольше всех;
- вытеснение случайное, т.е. удаляемая заявка выбирается случайным образом.

Отметим, что дисциплины буферизации актуальны только для систем с накопителями ограниченной емкости. В то же время в качестве моделей реальных систем часто используются СМО с накопителями неограниченной емкости, означающими, что любая поступившая заявка всегда найдет место в накопителе, т.е. не будет потеряна. И это несмотря на то, что в реальной системе емкость накопителя – ограничена. Такое предположение оправдано в тех случаях, когда вероятность переполнения ограниченной емкости в реальной системе мала. В частности, если вероятность переполнения накопителя ограниченной емкости меньше 10^{-4} , в качестве модели может использоваться СМО с накопителем неограниченной емкости, при этом погрешность результатов расчета характеристик функционирования не превысит 5%. Это означает, что в этом случае дисциплина буферизации практически не влияет на характеристики обслуживания заявок.

Дисциплины обслуживания (ДО), в зависимости от количества назначаемых на обслуживание заявок, могут быть разделены на дисциплины:

- одиночного режима;
- группового режима.

В **ДО одиночного режима** всякий раз на обслуживание назначается только одна заявка, причем просмотр очередей с целью назначения на обслуживание очередной заявки выполняется после завершения обслуживания каждой заявки.

В **ДО группового режима** всякий раз на обслуживание назначается группа заявок одной очереди, причем просмотр очередей с целью назначения на обслуживание выполняется только после завершения обслуживания всех заявок ранее назначенной группы. В предельном случае назначаемая на обслуживание группа заявок может включать в себя все заявки данной очереди. Заявки назначенной на обслуживание группы последовательно выбираются из очереди и обслуживаются устройством, после чего на обслуживание назначается следующая группа заявок другой очереди в соответствии с заданной ДО.

В реальных системах назначение заявок на обслуживание обычно связано с дополнительными затратами времени, называемыми накладными расходами. При групповом режиме накладные расходы меньше, чем при одиночном режиме, причем при увеличении размера группы, назначаемой на

обслуживание, уменьшение накладных расходов может оказаться значительным.

ДО группового режима широко применяются в беспроводных сетях связи под управлением протокола бесконфликтного доступа к среде и называются системами поллинга.

Система поллинга – объект, который содержит не менее двух накопителей и один или несколько серверов. В систему поллинга поступают заявки, которые необходимо обслужить (обработать). Заявки хранятся в накопителях и образуют очереди на обслуживание. Сервер по определённому правилу опрашивает накопители, подключается к одному из них и, в соответствии с выбранной дисциплиной обслуживания, последовательно обрабатывает заявки назначенной группы, называемой группой обработки.

В общем случае в системе могут использоваться ДО комбинированного режима, представляющего собой комбинацию одиночного и группового режимов.

В зависимости от отсутствия или наличия приоритетов между заявками разных классов дисциплины обслуживания делятся на два класса:

- бесприоритетные ДО;
- приоритетные ДО.

К бесприоритетным ДО относятся дисциплины обслуживания:

- в порядке поступления (FIFO – First In First Out);
- в обратном порядке (LIFO – Last In First Out);
- в случайном порядке (Random);
- в циклическом порядке (Cyclic), когда на обслуживание заявки выбираются в процессе циклического опроса накопителей в последовательности $1, 2, \dots, N$ (N – количество накопителей), после чего указанная последовательность повторяется.

Среди бесприоритетных ДО в технических системах наибольшее распространение получила дисциплина FIFO, которую в дальнейшем будем обозначать как ДО БП.

К приоритетным ДО относятся дисциплины обслуживания:

- с относительными приоритетами (ОП) – приоритеты учитываются только при выборе новой заявки на обслуживание, причем при поступлении заявки с более высоким приоритетом обслуживание низкоприоритетной заявки не прерывается;

- с абсолютными приоритетами (АП) – при поступлении заявки с высоким приоритетом прерывается обслуживание заявки с низким приоритетом и на обслуживание принимается поступившая высокоприоритетная заявка; при этом прерванная заявка может быть возвращена в накопитель или удалена из системы; если заявка возвращена в накопитель, то её дальнейшее обслуживание может быть продолжено с прерванного места или начато заново, то есть с начала;

- со смешанными приоритетами (СП), представляющими собой любую комбинацию БП, ОП и АП;

- с чередующимися приоритетами (ЧП) – приоритеты учитываются только в момент назначения на обслуживание новой группы (очереди) заявок после завершения обслуживания всех заявок предыдущей группы, при этом назначаемая на обслуживание группа может иметь произвольный размер (2 или более заявок) и в пределе включать в себя все заявки данной очереди;

- обслуживание по расписанию (ОР), когда заявки разных классов (находящиеся в разных накопителях) выбираются на обслуживание в соответствии с некоторым расписанием (планом), задающим последовательность опроса очередей заявок, например, в случае трех классов заявок (накопителей) расписание может иметь вид: {1, 2, 1, 3, 1, 2}.

Дисциплины БП, ОП, АП и СП относятся к дисциплинам одиночного режима, означающего, что после обслуживания каждой заявки выполняется поиск заявки с наивысшим приоритетом, поступившей в систему раньше других, которая и выбирается на обслуживание.

Дисциплины Сусліс, ЧП и ОР могут быть реализованы как ДО группового режима, если размер назначаемой на обслуживание группы больше 1, и как ДО одиночного режима, если размер назначаемой на обслуживание группы равен 1, при этом, например, ДО ЧП вырождается в ДО ОП.

Среди дисциплин обслуживания одиночного режима наибольший интерес представляют дисциплины со смешанными приоритетами (СП), обладающие общностью по отношению к дисциплинам БП, ОП и АП.

Для математического описания ДО СП используется **матрица приоритетов (МП)**, представляющая собой квадратную матрицу: $Q = [q_{ij} \mid i, j = 1, \dots, H]$, где H – число классов заявок, поступающих в систему.

Элемент q_{ij} матрицы задает приоритет заявок класса i по отношению к заявкам класса j и может принимать следующие значения:

- 0 – нет приоритета;
- 1 – приоритет относительный (ОП);
- 2 – приоритет абсолютный (АП).

Элементы МП должны удовлетворять следующим требованиям:

- $q_{ii} = 0$, так как между заявками одного и того же класса не могут быть установлены приоритеты;

- если $q_{ij} = 1$ или 2, то $q_{ji} = 0$, так как если заявки класса i имеют приоритет к заявкам класса j , то последние не могут иметь приоритет к заявкам класса i ($i, j = \overline{1, H}$).

Приоритеты заявок могут быть **статическими** и не изменяться в процессе функционирования системы или **динамическими**, т.е. изменяться в зависимости от разных факторов. Например, при достижении некоторого критического значения длины очереди заявок низкоприоритетного класса ему может быть предоставлен более высокий приоритет.

3.1.4. Режимы функционирования базовых моделей

Базовые модели могут отображать работу реальных систем в следующих режимах:

- **установившемся**, когда вероятностные характеристики системы не изменяются со временем;
- **неустановившемся**, когда характеристики системы изменяются со временем, что может быть обусловлено:
 - началом работы системы (**переходной режим**);
 - нестационарностью потоков заявок и длительностей обслуживания (**нестационарный режим**);
 - перегрузкой системы, когда интенсивность поступления заявок λ в систему с N устройствами превышает интенсивность их обслуживания $N\mu$, где $\mu = 1/b$ – интенсивность обслуживания заявок одним устройством; $\lambda > N\mu$ означает, что система не справляется с нагрузкой (**режим перегрузки**), при этом характеристики функционирования системы (в частности длина очереди перед устройством в системе с накопителем неограниченной емкости) растут неограниченно.

Обычно исследование систем проводится в предположении о существовании установившегося режима, непременным условием которого является требование отсутствия перегрузок. Это требование записывается в виде условия: $\lambda < N\mu$, т.е. нагрузка системы должна быть строго меньше единицы:

$$\rho = \min(y/N; 1) < 1, \quad (5)$$

где $y = \lambda/\mu$ – нагрузка системы.

Это условие для системы с одним устройством может быть записано в виде: $a > b$, где $a = 1/\lambda$ и $b = 1/\mu$ – средние значения интервала между поступающими в систему заявками и длительности обслуживания заявок в устройстве соответственно.

Следует обратить внимание на то, что речь идёт о среднем значении интервала между заявками и среднем значении длительности обслуживания. Если процессы поступления и обслуживания заявок детерминированные, то при отсутствии перегрузки очередь перед устройством не образуется. Такие системы, естественно, не представляют интереса и не рассматриваются в теории массового обслуживания. Очередь появится только в том случае, если процесс поступления заявок в систему и/или процесс обслуживания их в устройстве – случайные.

3.1.5. Характеристики систем с однородным потоком заявок

Рассмотрим характеристики функционирования систем с однородным потоком заявок, представляемых моделями с одним (рисунок 15) или с несколькими (рисунок 17) устройствами.

В систему с интенсивностью λ поступает случайный поток заявок, средняя длительность обслуживания которых в одном устройстве равна b .

Характеристики функционирования системы обычно рассчитываются для установившегося режима в предположении, что в системе отсутствуют перегрузки. Поскольку, в общем случае, процессы, протекающие в системе, носят случайный характер, то очевидно, что и характеристики функционирования системы представляют собой случайные величины, полное описание которых может быть представлено законами распределений в виде функции или плотности распределения. На практике обычно ограничиваются расчетом нескольких, чаще всего, двух первых моментов распределений случайных характеристик, что дает достаточно полное представление о свойствах исследуемой системы и позволяет эффективно решать задачи проектирования.

Ниже для простоты изложения материала в учебных целях и наглядности будем проводить все расчеты на уровне средних значений (математических ожиданий) характеристик функционирования, понимая, что при проектировании реальных систем на практике могут использоваться числовые моменты более высокого порядка, что приводит к применению сложных и громоздких математических зависимостей, но не меняет принципов проектирования такого рода систем.

В качестве основных характеристик базовых моделей с однородным потоком заявок используются следующие величины:

- **нагрузка системы:**

$$y = \lambda / \mu = \lambda b ; \quad (6)$$

нагрузка представляет собой интегральную характеристику, объединяющую два нагрузочных параметра: интенсивность поступления заявок в систему λ (частоту использования ресурса устройства) и среднюю длительность обслуживания заявки в устройстве b (количество потребляемого ресурса устройства при одном обращении);

- **вероятность потери заявок:**

$$\pi_n = \lim_{T \rightarrow \infty} \frac{N_n(T)}{N(T)}, \quad (7)$$

где T – время работы системы (наблюдения за системой); $N(T)$ – число заявок, поступивших в систему за время T ; $N_n(T)$ – число потерянных заявок (например, из-за ограниченной ёмкости накопителя или превышения допустимого времени ожидания) за время T ;

- **вероятность того, что заявка будет обслужена:**

$$\pi_0 = (1 - \pi_n) = \lim_{T \rightarrow \infty} \frac{N_0(T)}{N(T)}, \quad (8)$$

где $N_0(T)$ – число обслуженных (непотерянных) заявок за время T , причем $N_n(T) + N_0(T) = N(T)$ и $\pi_0 + \pi_n = 1$;

- **производительность системы**, измеряемая как интенсивность потока обслуженных заявок:

$$\lambda' = \pi_0 \lambda = (1 - \pi_n) \lambda; \quad (9)$$

для систем с накопителем неограниченной ёмкости, при условии отсутствия перегрузок, вероятность потери заявок $\pi_n = 0$ и, следовательно, производительность системы совпадает с интенсивностью поступления заявок в систему: $\lambda' = \lambda$;

- **интенсивность потока потерянных заявок** из-за ограниченной ёмкости накопителя:

$$\lambda'' = \pi_n \lambda = (1 - \pi_0) \lambda; \quad (10)$$

очевидно, что $\lambda' + \lambda'' = \lambda$;

- **коэффициент загрузки** или просто **загрузка** системы, представляющая собой долю времени, в течение которого система (в случае одноканальной СМО – устройство) работает, то есть выполняет обслуживание заявок:

$$\rho = \min\left(\frac{(1 - \pi_n) y}{N}; 1\right), \quad (11)$$

где N – число устройств в СМО, π_n – вероятность потери заявок, y – нагрузка системы; загрузка принимает значения в интервале $[0; 1]$ ($0 \leq \rho \leq 1$) в отличие от нагрузки, которая может принимать любые положительные значения ($y \geq 0$); в случае одноканальной СМО загрузка может трактоваться как вероятность того, что устройство работает (не простаивает);

- **коэффициент простоя системы:**

$$\pi = 1 - \rho;$$

коэффициент простоя показывает долю времени, в течение которого система простаивает, и может трактоваться как вероятность простоя;

- **среднее время ожидания** заявок в очереди: w ;

среднее время ожидания зависит от параметров системы и является одной из основных характеристик, на основе которой могут быть рассчитаны остальные характеристики с использованием фундаментальных зависимостей, поэтому часто задача исследования систем с очередями сводится, прежде всего, к определению значения времени ожидания;

- **среднее время пребывания** заявок в системе, складывающееся из времени ожидания w и времени обслуживания b : $u = w + b$;

- **средняя длина очереди** заявок:

$$l = \lambda w; \quad (12)$$

- **среднее число заявок в системе** (в очереди и на обслуживании в устройстве):

$$m = \lambda u. \quad (13)$$

Такие характеристики, как *время ожидания и пребывания заявок, длина очереди и число заявок в системе*, в общем случае, представляют собой случайные величины и полностью описываются соответствующими законами

распределений. На практике обычно ограничиваются расчетом указанных характеристик на уровне средних значений (математических ожиданий). Для более полного описания эффективности функционирования системы могут быть рассчитаны вторые моменты перечисленных характеристик (дисперсия или коэффициент вариации), показывающие их разброс относительно среднего значения.

Зависимости (12) и (13), связывающие средние значения временных (w , u) и безразмерных (l , m) характеристик, известны как **формулы Литтла** и представляют собой фундаментальные соотношения, справедливые для широкого класса моделей массового обслуживания.

Обычно исследование систем проводится в предположении о стационарности входящего потока заявок и длительности обслуживания. В этом случае условие существования установившегося режима совпадает с условием отсутствия перегрузок в системе и записывается в виде: $\rho < 1$.

3.1.6. Характеристики систем с неоднородным потоком заявок

Для базовой модели с неоднородным потоком заявок (рисунок 17), в которую поступают N классов заявок с интенсивностями $\lambda_1, \dots, \lambda_N$, а средние длительности обслуживания равны b_1, \dots, b_N , рассчитываются две группы характеристик обслуживания заявок:

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

Характеристики по каждому классу заявок $i = 1, \dots, N$ идентичны характеристикам базовой модели с однородным потоком (6) – (13) и в случае накопителей с неограниченной емкостью включают в себя:

- нагрузку, создаваемую заявками класса i : $y_i = \lambda_i / \mu_i = \lambda_i b_i$;
- загрузку системы заявками класса i : $\rho_i = \min(y_i / N; 1)$, где N - число устройств в СМО;
- среднее время ожидания заявок в очереди: w_i ;
- среднее время пребывания заявок в системе: $u_i = w_i + b_i$;
- среднюю длину очереди заявок: $l_i = \lambda_i w_i$;
- среднее число заявок в системе: $m_i = \lambda_i u_i$.

Характеристики объединенного (суммарного) потока заявок позволяют определить усредненные по всем классам заявок показатели эффективности функционирования системы, а именно:

- суммарную интенсивность поступления заявок в систему (интенсивность суммарного потока):

$$\Lambda = \sum_{i=1}^N \lambda_i; \quad (14)$$

- суммарную нагрузку Y и суммарную загрузку R системы:

$$Y = \sum_{i=1}^H y_i; \quad R = \min\left(\sum_{i=1}^H \rho_i; 1\right), \quad (15)$$

причем условие отсутствия перегрузок в системе с неоднородным потоком заявок имеет вид: $R < 1$;

- коэффициент простоя системы: $\pi = 1 - R$;
- среднее время ожидания W и среднее время пребывания U заявок объединенного потока в системе:

$$W = \sum_{i=1}^H \pi_i w_i; \quad U = \sum_{i=1}^H \pi_i u_i, \quad (16)$$

где $\pi_i = \lambda_i / \Lambda$ – коэффициент, учитывающий долю заявок класса i в суммарном потоке, который может трактоваться как вероятность того, что поступившая в систему заявка принадлежит классу i ;

- суммарная длина очереди и суммарное число заявок в системе:

$$L = \sum_{i=1}^H l_i; \quad M = \sum_{i=1}^H m_i. \quad (17)$$

Можно показать, что для характеристик объединенного потока справедливы те же фундаментальные соотношения, что и для однородного потока, а именно:

$$U = W + B; \quad L = \Lambda W; \quad M = \Lambda U, \quad (18)$$

где $B = \sum_{i=1}^H \pi_i b_i$ – среднее время обслуживания заявки объединенного потока.

3.2. Сетевые модели

Сетевые модели представляют собой сети массового обслуживания (СеМО), содержащие совокупность взаимосвязанных СМО, в среде которых циркулируют заявки. Сетевые модели могут быть изображены в виде **графа СеМО** (рисунок 18).

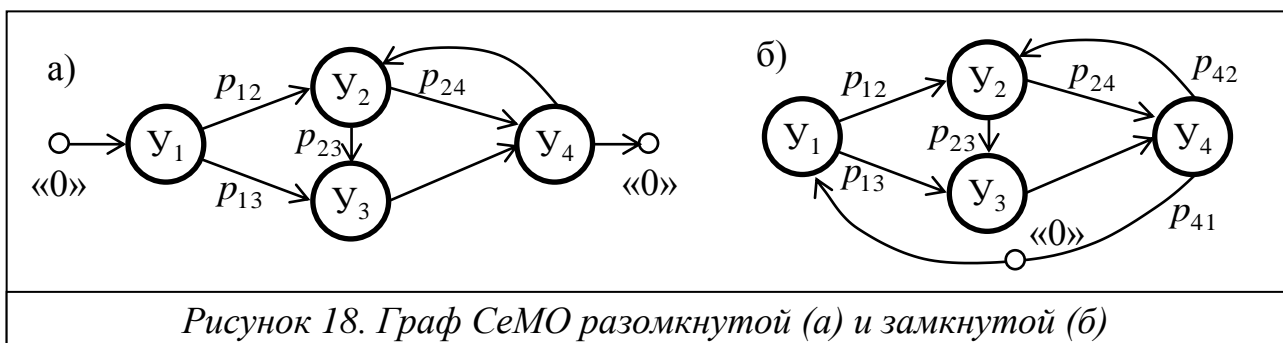


Рисунок 18. Граф СеМО разомкнутой (а) и замкнутой (б)

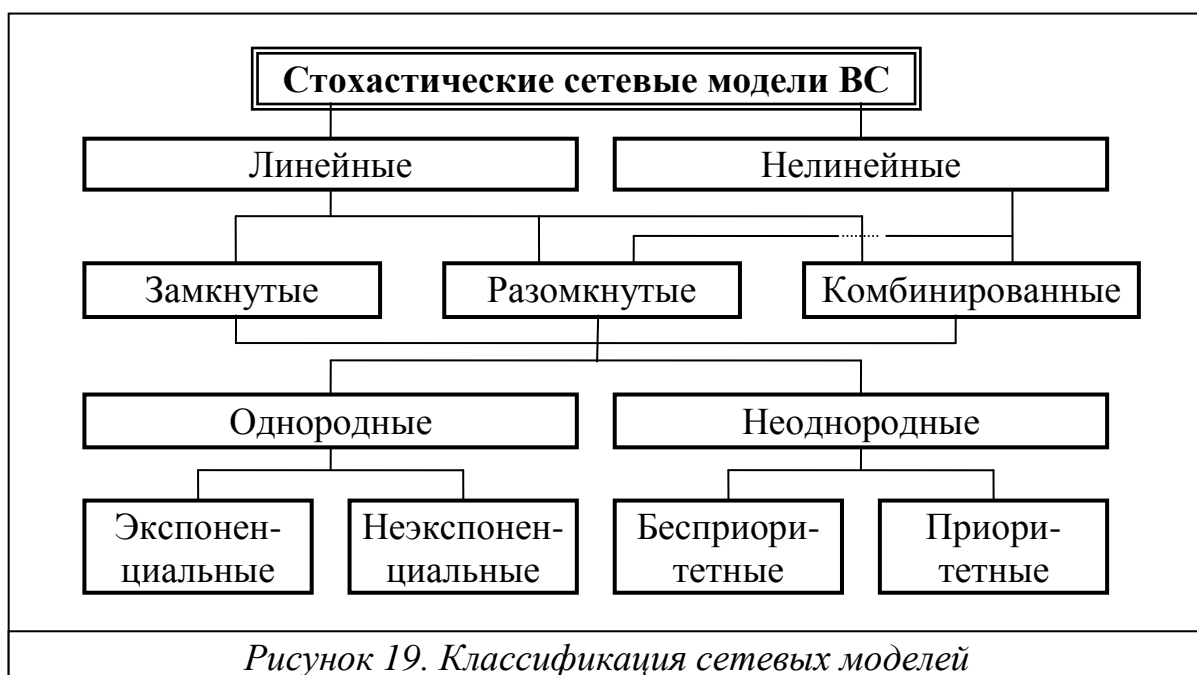
Основными элементами СеМО являются узлы (Y), представляющие собой СМО разных классов. В общем случае заявки поступают в сеть из внешней по отношению к СеМО среды с интенсивностью λ_0 и случайным образом перемещаются между узлами. По завершению обслуживания заявки в СеМО она возвращается во внешнюю среду. Путь движения заявок в СеМО

называется *маршрутом*, который описывается с помощью вероятностей передач: p_{ij} ($i, j = \overline{1, n}$) – вероятность того, что после обслуживания в узле i заявка перейдет в узел j , где n – количество узлов в СеМО.

3.2.1. Типы сетевых моделей

В зависимости от структуры и свойств исследуемых систем, их моделями могут служить СеМО различных классов (рисунок 19). Из всего множества различных сетевых моделей для решения задач системного анализа и проектирования будем в дальнейшем использовать стохастические линейные замкнутые и разомкнутые однородные СеМО.

Использование **стохастических моделей** обусловлено случайным характером процессов формирования нагрузки в реальных системах, в частности, процессов обработки и передачи данных в вычислительных системах. В терминах сетевой модели это означает, что интервалы времени между поступающими заявками и/или длительности их обслуживания в узлах представляют собой случайные величины, описываемые соответствующими законами распределений.



Линейность СеМО означает, что интенсивности поступления заявок в узлы сетевой модели связаны линейной зависимостью с интенсивностью поступления заявок в сеть (интенсивностью источника заявок) λ_0 : $\lambda_i = \alpha_i \lambda_0$ ($i = \overline{1, n}$), где n – количество узлов в сети. СеМО является линейной, если заявки в сети не теряются (например, из-за ограниченной емкости накопителя в узле) и не размножаются.

Принадлежность сетевой модели к классу разомкнутых или замкнутых сетевых моделей определяется числом циркулирующих в сети заявок.

В **разомкнутой (открытой) СеМО (РСеМО)** одновременно может находиться любое число заявок, в том числе и сколь угодно большое, т.е. от 0

до бесконечности. С РСемо связана внешняя среда, содержащая один или несколько внешних независимых источников заявок, которые генерируют заявки независимо от числа заявок, находящихся в сети. После завершения обслуживания в сети заявки возвращаются во внешнюю среду. Внешняя среда в РСемо обычно обозначается как дополнительный нулевой узел "0" (рисунок 18,а).

Замкнутая (закрытая) СеМО (ЗСеМО) не содержит внешних независимых источников заявок и характеризуется тем, что в ней циркулирует постоянное число заявок M . 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.

По аналогии с РСемо на выделенной дуге ЗСеМО отмечается условная точка "0" (рисунок 18,б), рассматриваемая как нулевой узел и трактуемая иногда как СеМО с нулевой длительностью обслуживания или как зависимый источник заявок, генерирующий заявки только в момент поступления некоторой заявки на вход источника. Выделение нулевого узла в ЗСеМО преследует двоякую цель: во-первых, достигается однозначность в представлении и математическом описании РСемо и ЗСеМО и, 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.

Однородность сетевой модели означает, что в СеМО циркулирует один класс заявок (однородный поток заявок), в отличие от неоднородного потока заявок, предполагающего наличие нескольких классов заявок.

Заявки в сетевой модели относятся к разным классам, если они различаются хотя бы одним из следующих факторов:

- длительностями обслуживания в узлах;
- приоритетами;
- маршрутами.

3.2.2. Параметры сетевых моделей

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

- число узлов в сети: n ;
- число обслуживающих устройств в узлах сети: K_1, \dots, K_n ;
- матрица вероятностей передач: $\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$, где p_{ij} – вероятность передачи заявки из узла i в узел j ;

- интенсивность λ_0 источника заявок, поступающих в РСeMO, или число заявок M , циркулирующих в ЗСеМО;

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

В случае неэкспоненциальных разомкнутых СеМО дополнительно необходимо задать законы распределения или, по крайней мере, вторые моменты распределения интервалов времени между заявками, поступающими в разомкнутую сеть, 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 $\mathbf{P}(h)$, интенсивности поступления заявок в РСeМО $\lambda_0(h)$ или число циркулирующих в ЗСеМО заявок $M(h)$, средние длительности обслуживания заявок в узлах сети $b_i(h)$ ($h = \overline{1, H}; i = \overline{1, n}$), а для приоритетных СеМО – дисциплины обслуживания заявок в узлах и дисциплины буферизации, если емкости накопителей ограничены. При необходимости могут быть заданы законы распределения интервалов между поступающими в РСeМО заявками и длительностей обслуживания заявок разных классов в узлах сети.

3.2.3. Характеристики сетевых моделей

Характеристики СеМО делятся на два класса:

- **узловые**, описывающие эффективность функционирования узлов;
- **сетевые**, описывающие функционирование СеМО в целом.

Состав узловых характеристик СеМО такой же, как и для базовых моделей (СМО) и для узла $j = \overline{1, n}$ включает в себя:

- нагрузку узла: $y_j = \lambda_j b_j = \alpha_j \lambda_0 b_j$, где α_j – коэффициент передачи, определяемый в процессе решения системы линейных алгебраических уравнений :

$$\lambda_j = \sum_{i=0}^n p_{ij} \lambda_i \quad (i = 0, 1, \dots, n); \quad (19)$$

- загрузку узла: $\rho_j = \min(y_j / K_j; 1)$, где K_j – число устройств в узле j ;
- среднее время ожидания заявок в очереди узла j : w_j ;
- среднее время пребывания заявок в узле j : $u_j = w_j + b_j$;
- среднюю длину очереди заявок в узле j : $l_j = \lambda_j w_j$;
- среднее число заявок в узле j : $m_j = \lambda_j u_j$.

В состав сетевых характеристик входят:

- среднее число заявок, ожидающих обслуживания в сети, определяемая как сумма длин очередей заявок по всем узлам:

$$L = \sum_{j=1}^n l_j, \quad (20)$$

где l_j – средняя длина очереди заявок в узле j ;

- среднее число заявок, находящихся в *разомкнутой* сети:

$$M = \sum_{j=1}^n m_j, \quad (21)$$

где m_j – среднее число заявок в узле j ;

в замкнутых СеМО число заявок M задано в качестве параметра и не может рассматриваться как характеристика сети, при этом выражение (20) может быть использовано для проверки правильности выполненных расчетов;

- среднее время ожидания заявок в сети:

$$W = \sum_{j=1}^n \alpha_j w_j, \quad (22)$$

где w_j – среднее время ожидания заявок в узле j ; α_j - коэффициент передачи для узла j , показывающий среднее число попаданий заявки в узел j за время ее нахождения в сети;

- среднее время пребывания заявок в сети:

$$U = \sum_{j=1}^n \alpha_j u_j, \quad (23)$$

где u_j – среднее время пребывания заявок в узле j ;

• производительность замкнутой СеМО λ_0 определяемая как интенсивность потока обслуженных заявок, проходящих через выделенный нулевой узел замкнутой сети:

$$\lambda_0 = \lambda_j / \alpha_j \quad (j = 1, \dots, n); \quad \lambda_0 = M / U. \quad (24)$$

Последняя формула представляет собой формулу Литтла для СеМО.

Для неоднородной СеМО перечисленные характеристики (20) – (24) определяются как для каждого класса в отдельности, так и для объединенного потока заявок.

3.2.4. Эквивалентные и толерантные преобразования сетевых моделей

Один и тот же объект, рассматриваемый на разных уровнях детализации, можно представить различными сетевыми моделями, характеристики которых одинаковы или отличаются на величину, не превосходящую заданной погрешности. При выполнении определенных условий такие модели легко преобразуются друг в друга. Две сетевые модели *эквивалентны*, если сравниваемые характеристики этих моделей не отличаются друг от друга. Две сетевые модели *толрантны* (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.

Эквивалентное преобразование разомкнутых сетей. Примером эквивалентного преобразования может служить преобразование разомкнутой однородной экспоненциальной сетевой модели в совокупность независимых СМО типа М/М/1, число которых равно количеству узлов РСемо. При этом в качестве параметров СМО $j = \overline{1, n}$ используются:

- интенсивность потока заявок, рассчитываемая через коэффициент передачи α_j : $\lambda_j = \alpha_j \lambda_0$;
- длительность обслуживания заявок, совпадающая с длительностью обслуживания в соответствующем узле Семо: b_j .

Характеристики СМО (среднее число заявок в очереди и системе, среднее время ожидания заявки в очереди и пребывания в системе, среднее число занятых устройств и т. д.) рассчитываются независимо друг от друга и являются узловыми характеристиками сети. Затем на их основе с использованием формул (20) – (24) определяются сетевые характеристики.

Толерантные преобразования. Определим условия, при которых ЗСемо могут быть представлены в виде РСемо. Такое преобразование позволяет пользоваться при расчете сетей более простыми математическими зависимостями. Для определения условий преобразования выявим особенности РСемо и ЗСемо и путь нивелировки этих особенностей.

РСемо состоит из n СМО и неограниченного источника с постоянной интенсивностью потока заявок, не зависящей от состояния сети, т.е. от числа заявок и их распределения в сети. Число заявок, находящихся в РСемо, – случайная величина, распределенная в диапазоне от 0 до бесконечности. Интенсивность поступающего в сеть потока заявок совместно с матрицей вероятностей передач однозначно определяет интенсивности входящих потоков в каждый узел сети. В результате, вероятности состояний каждого узла зависят только от его параметров (интенсивности входящего потока и среднего времени обслуживания), но не от параметров других узлов.

Замкнутая сеть состоит из n узлов и не содержит источника заявок неограниченной емкости, поэтому заранее не известны интенсивности потоков, входящих в каждый из узлов. Вероятности состояний отдельных узлов зависят от параметров всей сети в целом, т. е. от параметров других узлов сети. Распределение заявок между узлами сети случайно, но их общее число M постоянно. В такой сети источником может считаться любой узел, но емкость каждого источника, т. е. число заявок в нем, ограничена. Так как параметры узлов сети различны, то в одних узлах будет скапливаться больше заявок, а в других меньше. Наибольшее число заявок будет скапливаться в наиболее загруженном узле. Поэтому такой узел – узкое место сети.

Пусть узел j , содержащий K_j устройств со средним временем обслуживания b_j в каждом из них, является узким местом сети. Чем больше нагрузка узла j , тем большая доля всех заявок M скапливается в нем и тем выше вероятность того, что все устройства данного узла заняты обслуживанием заявок. Можно показать, что при бесконечно большом числе заявок M в

многоканальном узле, являющимся узким местом сети, даже при неограниченно большом числе устройств скапливается с единичной вероятностью большое число заявок. Загрузка каждого из устройств данного узла стремится к единице, а интенсивность потока на его выходе – к величине K_j/b_j . Следовательно, замкнутая сеть становится подобной разомкнутой сети, в которой источником с постоянной интенсивностью K_j/b_j является узел j . Таким образом, узел j замкнутой сети будет потенциальным источником, если его загрузка максимальна. Учитывая, что $\rho_j = \lambda_j b_j / K_j$ и $\lambda_j = \alpha_j \lambda_0$, получим условие толерантности замкнутой и разомкнутой сети:

$$\rho_j = \alpha_j b_j \lambda_0 / K_j = \max_{i=1, \dots, n} (\alpha_i b_i \lambda_0 / K_i) = 1 \quad (j=1, \dots, n).$$

Если в замкнутой сети имеется несколько узлов, для которых выполняется данное условие, то сеть также преобразуется в разомкнутую. Интенсивности потоков в узлах рассчитываются, как и при одном источнике, на основе равенства $\lambda_j = \alpha_j \lambda_0$.

Часто оказывается, что даже при конечном числе заявок M замкнутую сеть можно представить как разомкнутую. Условием, при котором замкнутая сеть может рассматриваться как разомкнутая, т. е. условием толерантности замкнутой и разомкнутой сети является выполнение неравенства $\max(\rho_j / \rho_i) \gg 1$ для всех $i=1, \dots, n$, кроме $i=j$.

Индекс j выделяет наиболее загруженную систему сети, а соотношение $x \gg y$ означает, что x по крайней мере на порядок больше y . При достаточно большом числе заявок M в сети это условие значительно ослабляется. Уже при $M > 20$ замкнутую сеть можно рассматривать как разомкнутую, даже при условии: $\max(\rho_j / \rho_i) > 1$.

4. Модели вычислительных систем

4.1. Принципы структурно-функциональной организации вычислительных систем

В качестве объектов проектирования и исследования в теории вычислительных систем рассматриваются вычислительные системы разных классов. Ниже вводятся понятия, используемые для систематизации и выявления классов вычислительных систем, характеризующихся одинаковыми или близкими свойствами, позволяющими унифицировать процесс исследования и проектирования различных систем *на системном уровне*.

4.1.1. Основные понятия теории вычислительных систем

Вычислительная система (ВС) – совокупность *технических* и *программных* средств, ориентированных на решение задач, связанных с обработкой данных и получением информации в виде результата решения задачи.

К техническим средствам ВС относятся устройства ввода-вывода, хранения и обработки данных, образующие *электронную вычислительную*

машину (ЭВМ, компьютер). Другими словами, ЭВМ или просто вычислительная машина (ВМ) представляет собой совокупность взаимосвязанных устройств для организации ввода, хранения, автоматической обработки по заданной программе и вывода данных. Такими устройствами являются: центральный процессор (ЦП), оперативная (основная) память (ОП), внешние устройства (ВУ), включающие устройства ввода-вывода (УВВ) и внешние запоминающие устройства (ВЗУ), процессоры (каналы) ввода-вывода (ПВВ, КВВ).

Вычислительные системы могут быть построены на основе однопроцессорных вычислительных машин или многомашинных и многопроцессорных вычислительных комплексов.

Вычислительный комплекс (ВК) – совокупность технических средств, содержащая несколько центральных процессоров. ВК может быть реализован в виде одной вычислительной машины с несколькими ЦП (МПВК – многопроцессорный ВК) или путём объединения нескольких однопроцессорных вычислительных машин, расположенных на небольшом расстоянии друг от друга (ММВК – многомашинный ВК).

Основной целью построения ВК является обеспечение высокой надежности и/или производительности, не достижимой для однопроцессорных вычислительных машин.

Программные средства ВС делятся на два класса:

- **системное программное обеспечение**, включающее операционную систему (ОС), основными составляющими которой, с точки зрения организации эффективного функционирования ВС, являются *управляющие программы* (УП);
- **прикладное программное обеспечение** в виде множества *прикладных программ* (ПП), обеспечивающих ориентацию ВС на решение задач конкретной области.

На системном уровне в качестве основной единицы работы в ВС рассматривается задача, представляющая совокупность определенной прикладной программы с определенным набором данных. Причиной инициализации задачи может быть задание, команда, запрос или транзакция.

Выполнение задач в ВС называется вычислительным процессом.

Последовательность решения задач в ВС, т.е. управление вычислительным процессом, определяется управляющими программами операционной системы.

К программным средствам ВС также относятся системы управления базами данных (СУБД), которые совместно с базами данных (БД) образуют информационное обеспечение.

Вычислительная сеть или, точнее, компьютерная сеть (сеть ЭВМ) на системном уровне рассматривается как совокупность вычислительных систем разных классов, расположенных, в общем случае, на значительных расстояниях друг от друга и объединенных средой передачи данных, образующей телекоммуникационную сеть. Компьютерная сеть реализует две основные функции:

- обработку данных;

- передачу данных.

Понятие «вычислительная система» в широком смысле слова включает в себя и компьютерные сети, которые могут рассматриваться как системы более высокого уровня.

В заключение ответим на вопрос: в чем различие между понятиями «данные» и «информация»?

Данные – множество сведений, представленных в виде, пригодном для обработки с целью получения какого-нибудь вывода, решения. Количественной мерой данных является *объём* – количество единиц данных, измеренных в байтах, словах, страницах и т.п.

Информация – сведения, полученные с определенной целью путем обработки данных. Количественной мерой информации является энтропия – мера неопределенности информации.

Из этих определений следует, что данные – это любое множество сведений, а информация – это сведения, полученные с определенной целью и несущие в себе новые знания для того, кто эту информацию получает.

4.1.2. Структурная организация вычислительных систем

Структурная организация ВС определяется:

- составом (номенклатурой и количеством) устройств, входящих в состав системы;
- способом связи устройств между собой;
- параметрами устройств.

На системном уровне ВС обычно рассматривают как совокупность подсистемы ЦП-ОП, являющейся центральной в составе ВС, и подсистемы ввода-вывода. Каноническая структура однопроцессорной ВС представлена на рисунке 20.

Центральный процессор (ЦП) реализует обработку данных и управление вычислительным процессом.

Оперативная память (ОП) используется для хранения программ и данных, участвующих в вычислительном процессе и в любой момент времени доступных центральному процессору.

ЦП и ОП образуют подсистему ЦП-ОП (ядро ВС).

Процессоры (каналы) ввода-вывода (ПВВ, КВВ) представляют собой специализированные процессоры с ограниченными функциональными возможностями, необходимыми для организации обмена данными между ОП и ВУ без участия ЦП, что позволяет

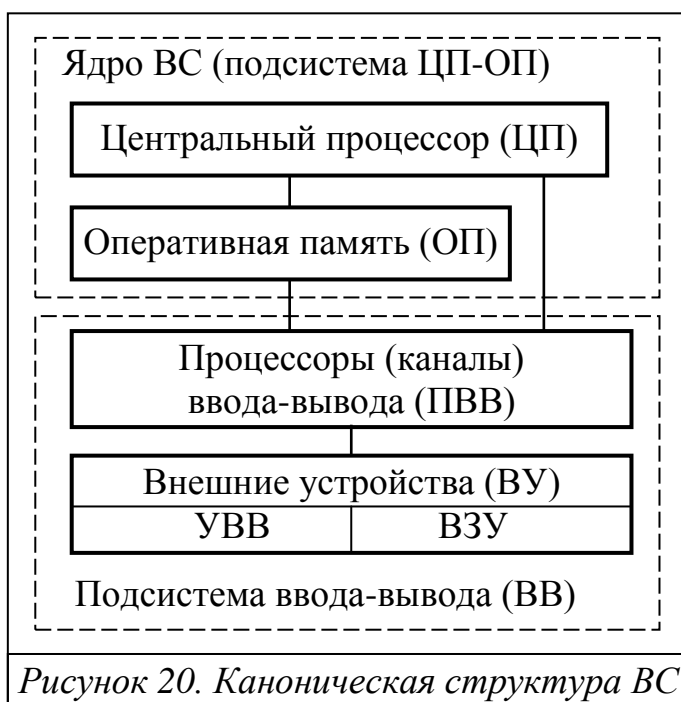


Рисунок 20. Каноническая структура ВС

существенно повысить эффективность функционирования ВС. В персональных компьютерах функции обмена с ВУ реализуются контроллерами внешних устройств.

Внешние устройства (ВУ) представляют собой периферийное оборудование ВМ и включают в себя:

- быстрые устройства – ВЗУ: накопители на магнитных и оптических дисках и магнитных лентах, предназначенные для длительного хранения программ и данных;
- медленные устройства – УВВ: различного рода терминалы, дисплеи, устройства печати и т.п., предназначенные для организации обмена данными с внешней средой (пользователями).

ПВВ и ВУ образуют подсистему ввода-вывода.

Комплексирование двух и более ВМ, входящих в состав вычислительного комплекса, может выполняться на разных уровнях.

Наиболее существенен в структурной организации ВК способ связи между ЦП и оперативной памятью комплекса, в зависимости от которого ВК разделяются на комплексы:

- с индивидуальной (раздельной) памятью – *многомашинные вычислительные комплексы (ММВК)*;
- с общей (полнодоступной) памятью – *многопроцессорные вычислительные комплексы (МПВК)*.

ММВК представляют собой совокупность взаимосвязанных ВМ, а МПВК – одну ВМ с несколькими процессорами.

На рисунке 21 представлены канонические структуры ММВК и МПВК.

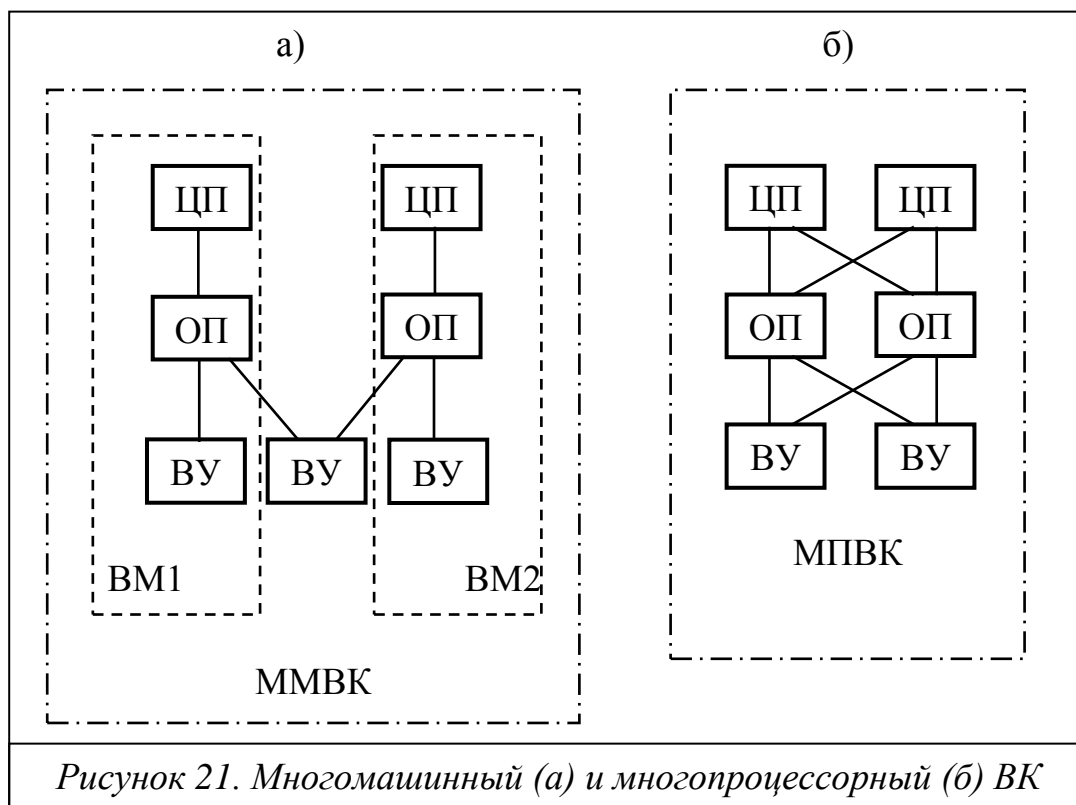


Рисунок 21. Многомашинный (а) и многопроцессорный (б) ВК

МПВК характеризуются тем, что любой процессор ВК имеет непосредственный доступ к любому модулю ОП. Более того, в МПВК обычно существует связь между любой парой устройств (рисунок 21,б).

Структурная организация *компьютерных сетей* определяется:

- составом узлов (номенклатура и количество сетевых устройств, ВМ, ВС и терминалов) и топологией сети передачи данных (СПД);
- производительностью узлов связи и пропускной способностью каналов связи.

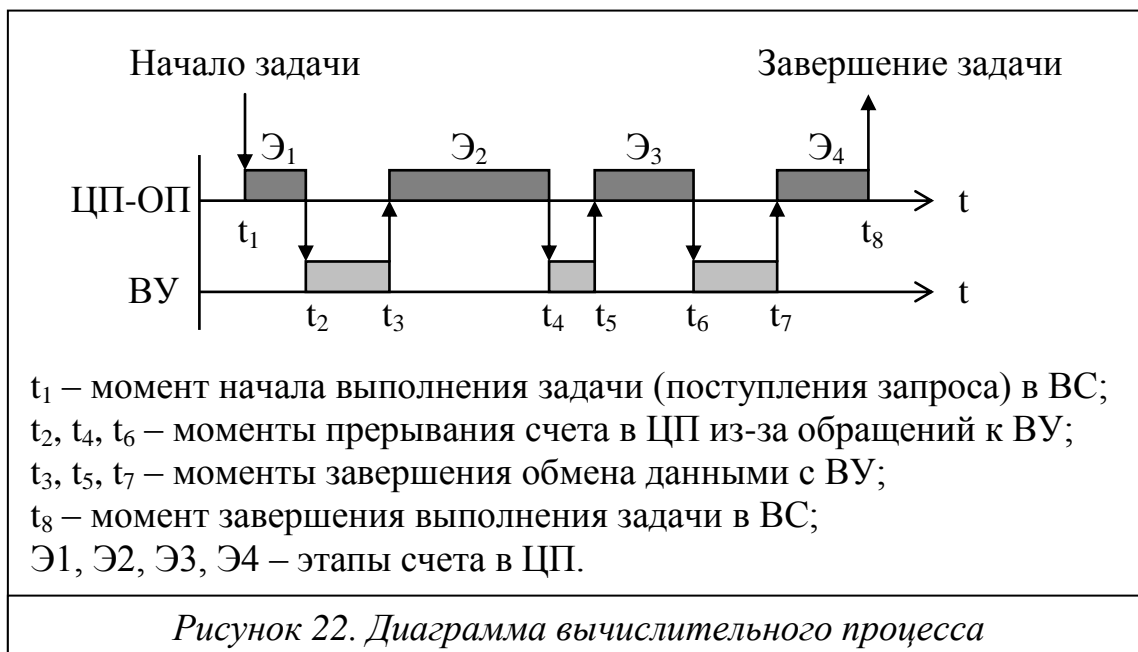
4.1.3. Функциональная организация вычислительных систем

Функциональная организация ВС – способ реализации возложенных на систему функций по организации процессов обработки и передачи данных, реализуемых в виде определенных режимов функционирования ВС.

Вычислительный процесс на системном уровне можно представить как последовательное чередование двух этапов (рисунок 22):

- 1) процессорная обработка (выполнение команд в ЦП);
- 2) ввод-вывод данных (обмен с внешними устройствами).

Процессорная обработка на системном уровне рассматривается как последовательность непрерывных интервалов (этапов счета Э1, Э2, ...) выполнения команд в ЦП в процессе решения задачи. Завершение этапа счета связано с обращением к ВУ или с завершением задачи. Количество этапов счета на единицу больше, чем число обращений к ВУ, поскольку выполнение задачи начинается и завершается всегда в ЦП.



Средняя длительность этапа счета определяется ресурсоемкостью θ_0 , измеряемой количеством команд (инструкций), выполняемых в процессе этапа счета, и производительностью (быстродействием) ЦП $V_{ЦП}$:

$$t_0 = \theta_0 / V_{ЦП}. \quad (25)$$

При рассмотрении процессорной обработки следует различать ресурсоемкость задачи θ и ресурсоемкость этапа счета θ_0 :

$$\theta_0 = \theta / (D_{\text{ВУ}} + 1), \quad (26)$$

где $D_{\text{ВУ}}$ – число обращений к ВУ в процессе выполнения задачи.

Если в процессе выполнения задачи нет обращений к внешним устройствам ($D_{\text{ВУ}} = 0$), например в системах реального времени, то ресурсоемкость задачи и ресурсоемкость этапа счета совпадают: $\theta_0 = \theta$.

Состав параметров, описывающих *процесс ввода-вывода* данных, зависит от типа внешнего устройства и режима его функционирования. Например, при обмене с накопителями на магнитных дисках могут использоваться различные механизмы (дисциплины) доступа к данным, уменьшающие время доступа. Кроме того, эффективность процесса ввода-вывода может быть увеличена за счет механизма кэширования.

Существенное влияние на характеристики функционирования ВС, такие как время ответа и производительность, оказывает режим *функционирования системы*.

В зависимости от количества задач, одновременно выполняемых в системе, различают:

- однопрограммный (однозадачный) режим функционирования ВС, при котором в системе выполняется только одна программа (задача);
- мультипрограммный (многозадачный) режим функционирования ВС, при котором в системе одновременно выполняется несколько задач, причем соответствующие им программы загружены в ОП полностью или частично.

Количество задач, одновременно находящихся в системе, называется *уровнем мультипрограммирования*.

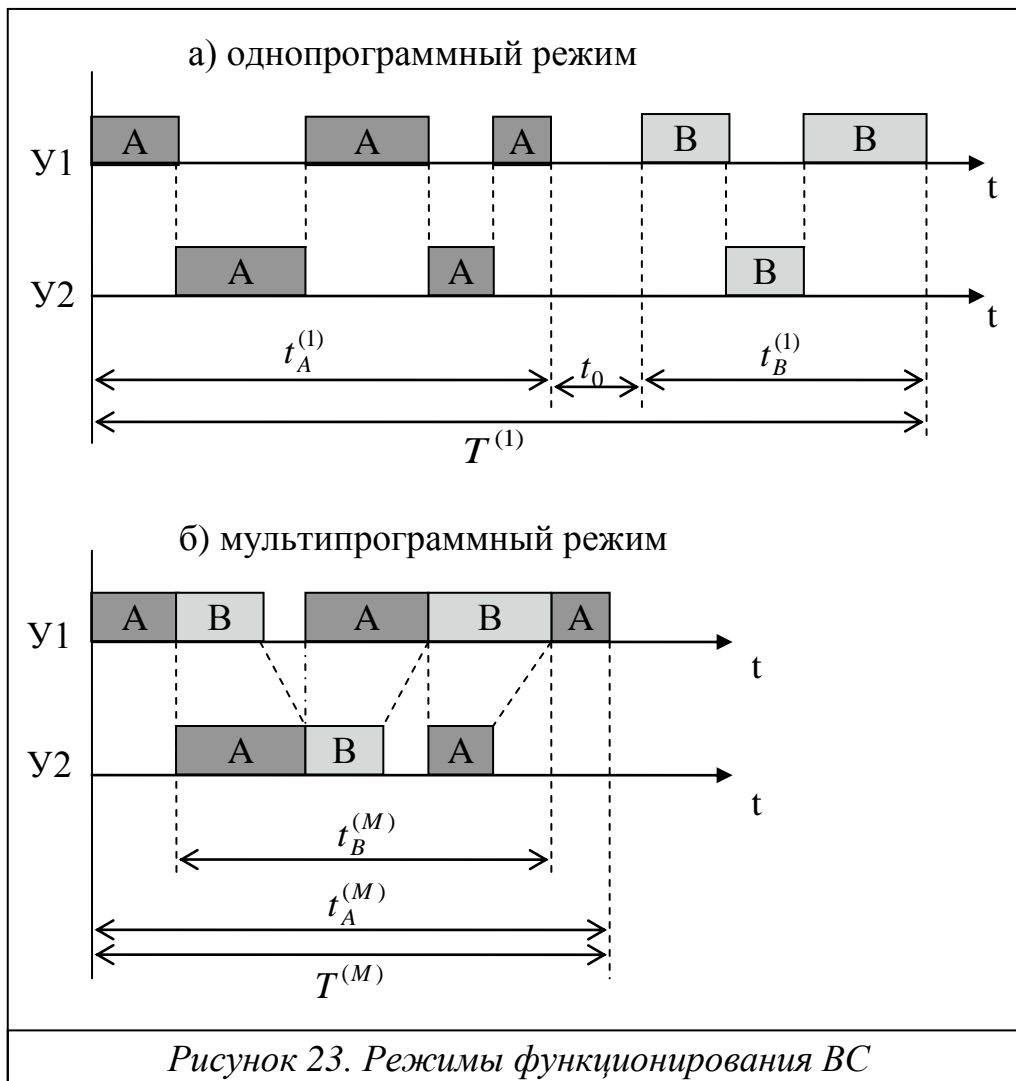
Основная цель мультипрограммирования – *увеличение системной производительности* ВС за счет *увеличения загрузки* устройств системы, достигаемой в результате *распараллеливания* работы устройств.

Эффект применения мультипрограммного режима иллюстрируется на рисунке 23. В ВС с двумя устройствами У1 и У2 выполняются две задачи А и В в однопрограммном (рисунок 23,а) и мультипрограммном (рисунок 23,б) режимах.

Из представленных диаграмм видно, что:

1) системная производительность ВС, измеряемая количеством задач выполненных за единицу времени, при мультипрограммном режиме $\lambda_0^{(M)} = 2/T^{(M)}$ выше, чем при однопрограммном $\lambda_0^{(1)} = 2/T^{(1)}$: $\lambda_0^{(M)} > \lambda_0^{(1)}$, так как $T^{(M)} < T^{(1)}$;

2) среднее время ответа (время пребывания задачи в ВС) при однопрограммном режиме $u^{(1)} = (t_A^{(1)} + t_B^{(1)})/2$ меньше, чем при мультипрограммном режиме $u^{(M)} = (t_A^{(M)} + t_B^{(M)})/2$: $u^{(1)} < u^{(M)}$, что при мультипрограммном режиме обусловлено задержками, связанными с ожиданием перед занятыми устройствами.



Таким образом, мультипрограммирование позволяет увеличить производительность ВС за счет более полного и эффективного использования устройств системы, однако оперативность при этом уменьшается. Отметим, что наибольшая оперативность ВС достигается при однопрограммном режиме.

В зависимости от назначения ВС, формулируемого в виде требований к качеству функционирования системы, вычислительный процесс может быть реализован тремя способами:

- 1) пакетная обработка;
- 2) оперативная обработка;
- 3) реальный масштаб времени.

Пакетная обработка направлена на повышение системной производительности ВС, измеряемой числом задач, решаемых системой за единицу времени, и, как следствие этого, на снижение стоимости обработки данных за счет эффективного использования ресурсов системы.

Пакетная обработка характеризуется:

- большим объемом вводимых-выводимых данных и вычислений, приходящихся на одну задачу;
- низкой интенсивностью взаимодействия пользователей с ВС ($\sim 10^{-1} - 10^0$ взаимодействий в час).

Основной принцип планирования пакетной обработки – *максимальная загрузка всех устройств ВС* или, что то же самое, обеспечение *минимального времени решения всех задач* пакета.

Пакетная обработка характерна для больших ЭВМ (мэйнфреймов) и суперЭВМ, которые решают задачи, характеризующиеся большой ресурсоемкостью.

Оперативная обработка направлена на уменьшение времени ответа и характеризуется:

- малым объемом вводимых-выводимых данных и вычислений, приходящихся на одну задачу;
- высокой интенсивностью взаимодействия пользователей с системой ($\sim 10^1 - 10^2$ взаимодействий в час).

Основной принцип планирования оперативной обработки – *короткая задача вперед*, т.е. задача, имеющая меньшую ресурсоемкость, должна обрабатываться в первую очередь.

Оперативная обработка характерна для информационно-справочных систем, систем резервирования билетов и т.п.

В ВС общего назначения, когда ресурсоемкость задач заранее неизвестна, этот принцип реализуется путем разделения между выполняемыми задачами общего для всех задач ресурса – процессора. При этом каждой задаче выделяется квант времени длительностью порядка (0,1 – 1) секунд, в течение которого задача выполняется в процессоре. По истечении этого кванта, если задача не завершилась, она прерывается и переходит в состояние ожидания, а процессор начинает выполнять следующую задачу (рисунок 24). Такой режим обработки задач называется режимом разделения времени или режимом квантования.

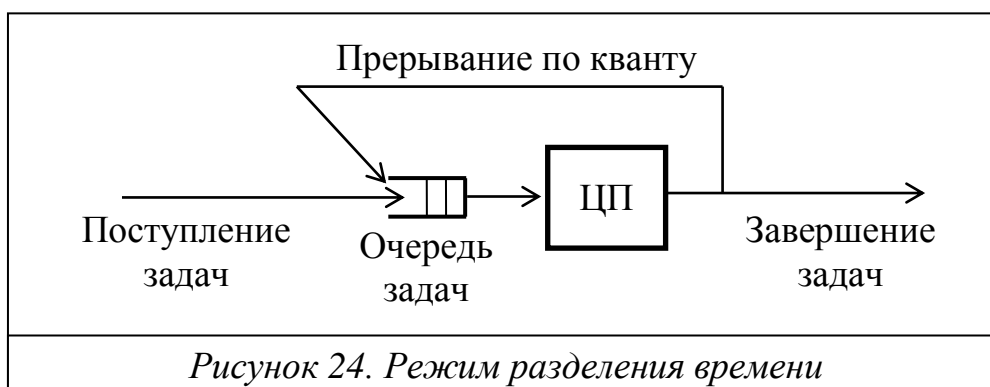


Рисунок 24. Режим разделения времени

В ВС, предназначенных для управления реальными объектами и процессами, выдача управляющих воздействий осуществляется в темпе, диктуемом внешним (управляемым) объектом. Такой режим функционирования системы называется реальным масштабом времени (РМВ) или просто режимом реального времени, а системы, работающие в РМВ, - системами реального времени (СРВ) или информационно-управляющими системами (ИУС).

РМВ характеризуется:

- малым объемом вводимых-выводимых данных, представляемых в виде запросов на обработку;
- высокой интенсивностью взаимодействия внешнего объекта с системой (более 10^3 взаимодействий в час);
- наличием жестких ограничений на время пребывания запросов в системе, называемое временем реакции.

Понятие "жесткое ограничение" подразумевает два обстоятельства:

- 1) *малые значения времени реакции*, измеряемые долями секунд;
- 2) *жесткие требования по выполнению заданных ограничений*, превышение которых может привести к резкому ухудшению функционирования системы или выходу системы из строя.

РМВ обеспечивается за счет рационального распределения приоритетов между задачами, определяющих последовательность их выполнения, или за счет формирования эффективного расписания выполнения задач.

Функциональная организация компьютерной сети определяется:

- 1) способом организации каналов связи между взаимодействующими абонентами и способом передачи данных по каналам (коммутация каналов, коммутация сообщений, коммутация пакетов);
- 2) алгоритмом выбора маршрутов передачи данных (статические или фиксированные; динамические или адаптивные);
- 3) способом управления потоками данных в сети (механизмы тайм-аута и скользящего окна, распределение приоритетов, ограничение доступа в сеть).

4.2. Параметры вычислительных систем

Параметры вычислительных систем можно разбить на две группы:

- **структурно-функциональные параметры**, описывающие:
 - состав и структуру ВС (*структурные параметры*): типы и взаимосвязь устройств и их технические данные (быстродействие ЦП, емкости ОП и ВЗУ, пропускные способности каналов связи и т.п.), количество узлов, входящих в состав сети, и их взаимосвязь (топология сети) и т.д.;
 - режим функционирования ВС (*функциональные параметры*): стратегии управления обработкой данных (последовательность выполнения прикладных задач, приоритеты задач и т.п.) и передач данных в компьютерной сети (способ коммутации, метод доступа к каналу связи, алгоритм выбора маршрута передачи данных в сети);
 - **нагрузочные параметры**, описывающие взаимодействие ВС с внешней средой, т.е. нагрузку, создаваемую в сети решаемыми прикладными задачами и передаваемыми в вычислительной сети данными: число типов задач, ресурсоемкость каждой задачи, объем занимаемой памяти, длина передаваемых по сети сообщений и т.п.

4.2.1. Параметры структурно-функциональной организации ВС

Структурные параметры ВС описывают:

- номенклатуру и количество устройств каждого типа, входящих в состав системы;

- способ связи устройств между собой;
- технические данные устройств, которые при решении задач системотехнического проектирования ВС трактуются как параметры.

При разработке моделей функционирования ВС на системном уровне в качестве элементов, параметры которых должны быть заданы, обычно рассматривают:

- центральный процессор (ЦП);
- оперативную память (ОП);
- процессоры (каналы) ввода-вывода (ПВВ, КВВ);
- внешние запоминающие устройства (ВЗУ)
- устройства ввода-вывода (УВВ).

В качестве параметров ЦП на системном уровне обычно используется среднее время выполнения команды [мкс, нс,...] или быстродействие (производительность) процессора, задаваемое в зависимости от класса ВС в виде следующих величин:

- MIPS (Million Instructions Per Second) – единица измерения быстродействия, равная одному миллиону инструкций (команд) в секунду;
- FLOPS (или flops или flop/s) (Floating point Operations Per Second) — величина, используемая для измерения производительности суперЭВМ и показывающая, число операций с плавающей запятой, выполняемых за секунду.

В качестве параметров ОП на системном уровне используются:

- ёмкость ОП [Мбайт, Гбайт, ...];
- время обращения (доступа) к ОП [мкс, нс,...].

В качестве параметра ПВВ используется пропускная способность [Мбайт/с, Гбайт/с, ...].

В качестве параметров ВЗУ на системном уровне используются следующие величины:

- емкость ВЗУ [Гбайт, Тбайт,...];
- скорость обмена [Мбайт/с, ...];
- скорость вращения НМД [тыс. оборотов в минуту] или скорость перемотки магнитной ленты в НМЛ [метров в секунду];
- среднее время доступа к данным, называемое также временем поиска данных, затрачиваемое на перемещение головки или блока головок чтения-записи на требуемую дорожку (цилиндр) [мс, мкс].

В качестве параметра УВВ на системном уровне может использоваться скорость ввода-вывода [символов или строк в секунду].

Параметры структурной организации *компьютерных сетей* можно разбить на три группы:

- параметры топологии сети: количество и типы узлов связи и их взаимосвязь, задаваемая с помощью матрицы связей (инцидентности);
- параметры узлов, представляющих собой, в общем случае, средства вычислительной техники, реализующие функции коммутации и маршрутизации

в соответствии с принятой в данной сети технологией передачи данных; таким образом, в качестве системных параметров узлов связи могут использоваться те же параметры, что и для вычислительных систем;

- параметры каналов связи: тип канала (симплексный, дуплексный, полудуплексный), стоимостные параметры (удельная стоимость, стоимостная зависимость от пропускной способности), а также основной параметр – пропускная способность канала связи, определяемая как количество данных, измеряемое в битах, которое может быть передано по каналу за единицу времени (обычно за секунду): кбит/с, Мбит/с, Гбит/с,

Совокупность параметров, описывающих функциональную организацию ВС, определяется способами управления вычислительным процессом и процессом передачи данных, задаваемых в виде дисциплин буферизации и обслуживания, распределения приоритетов и т.п. Конкретный состав функциональных параметров определяется целью исследований и видом модели, используемой при проектировании ВС.

4.2.2. Параметры нагрузки ВС

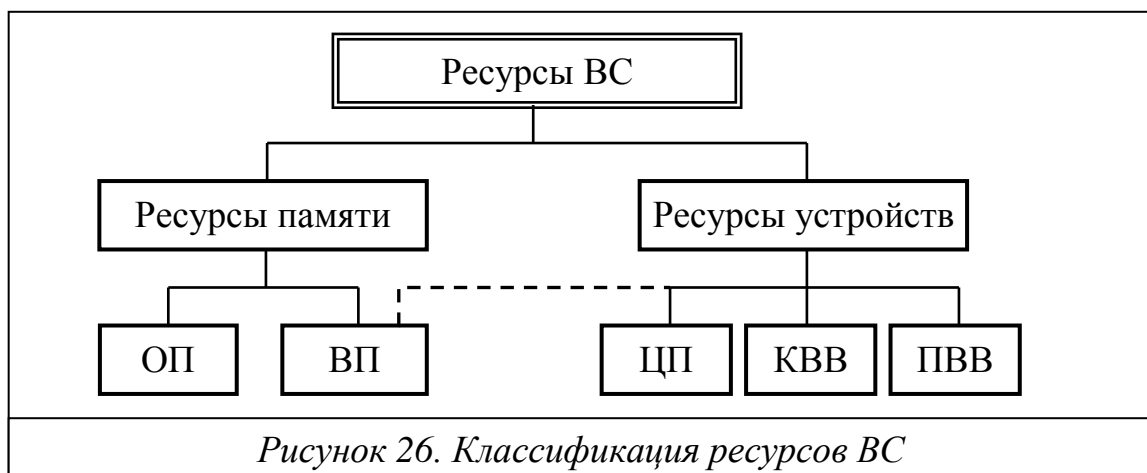
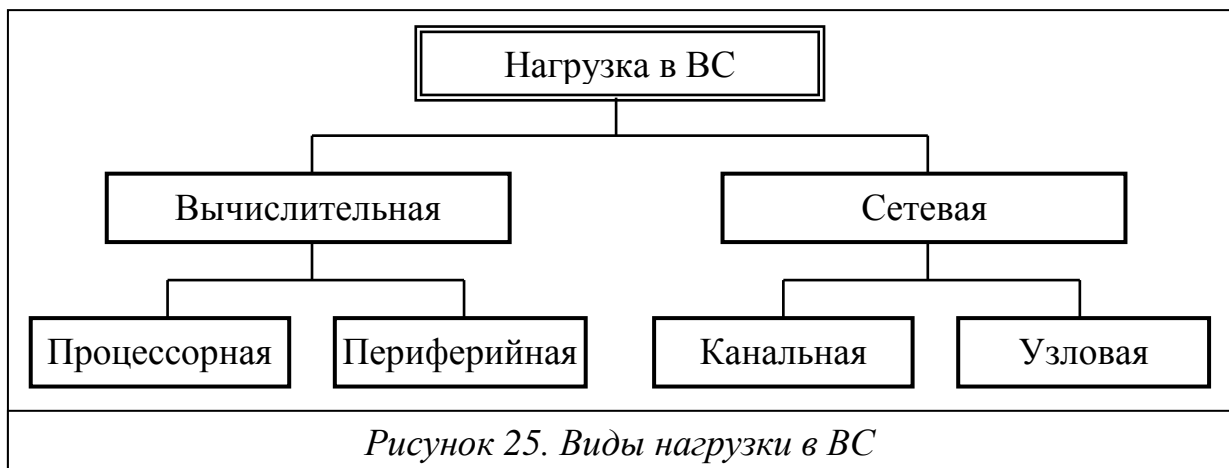
Среди перечисленных выше трех групп параметров ВС наибольший интерес представляют нагрузочные параметры. Этот интерес обусловлен несколькими причинами, в частности, сложностью определения и задания их значений, которые в принципе могут быть получены путем измерений на реальных вычислительных системах и сетях с использованием специальных программных (а иногда и аппаратных) мониторов. Очевидно, что это оказывается невозможным для еще не существующих, а только проектируемых систем и сетей. В этом случае нагрузочные параметры могут быть заданы априорно, либо проектирование ведется в расчете на некоторую ожидаемую, возможно предельную, нагрузку.

Нагрузка в вычислительной системе (сети) определяется потребностями задач и передаваемых данных (сообщений) в ресурсах системы и задается совокупностью *нагрузочных параметров*. Нагрузка, создаваемая в вычислительных системах и сетях, порождается *источниками нагрузки*, в качестве которых могут выступать пользователи, формирующие задачи для выполнения в ВС или сообщения, передаваемые в компьютерной сети.

В зависимости от типа используемых ресурсов и формы обработки данных различают следующие виды нагрузки в ВС (рисунок 25):

- вычислительная, которая делится на:
 - процессорную, создаваемую в ядре ВС – подсистеме ЦП-ОП;
 - периферийную, создаваемую во внешних устройствах, прежде всего в ВЗУ;
- сетевая, которая делится на:
 - канальную, создаваемую в каналах связи компьютерной сети;
 - узловую, создаваемую в маршрутизаторах и коммутаторах.

Ресурсы ВС делятся на ресурсы памяти и ресурсы устройств (рисунок 26).



Ресурс памяти – это емкость, которая может быть распределена в пространстве между несколькими задачами. Ресурсы памяти делятся на ресурсы оперативной и внешней памяти.

Ресурс устройства – это количество работы, которую может выполнить устройство.

Устройство характеризуется *производительностью* (быстродействием, скоростью работы) и определенное количество работы выполняет за вполне определенное время. Поэтому в качестве ресурса устройства часто рассматривают *время*, в течение которого устройство выполняет данную работу. Поскольку любое устройство ВС в каждый момент времени может выполнять только одну работу (задачу), то говорят, что ресурс устройства распределяется во времени. К устройствам, ресурсы которых распределяются только во времени, относятся процессоры, каналы и устройства ввода-вывода.

Внешние запоминающие устройства распределяются как в пространстве, так и во времени.

Ресурсоемкость – потребность задачи в определенном ресурсе (памяти или устройства). Ресурсоемкость задается матрицей ресурсоемкости задач: $X = [x_{ij} \ (i = \overline{1, M}; j = \overline{1, N})]$, где x_{ij} – потребность задачи Z_i в ресурсе Y_j .

Если Y_j является памятью, то значение x_{ij} определяет количество единиц информации (в битах или байтах), которые должны быть размещены в памяти

Y_j при выполнении задачи Z_i . Если Y_j – устройство, то x_{ij} – время, необходимое для выполнения задачи Z_i на этом устройстве, которое может быть определено через количество работы A_{ij} , выполняемой задачей Z_i в устройстве Y_j , и производительность V_j устройства Y_j : $x_{ij} = A_{ij}/V_j$. Если задача Z_i не использует ресурс Y_j , то $x_{ij} = 0$.

Источники нагрузки могут быть двух типов:

- внешние, порождающие нагрузку во внешней по отношению к рассматриваемой системе среде;
- внутренние, порождающие нагрузку внутри рассматриваемой системы.

К внешним источникам нагрузки относятся:

- пользователи, формирующие запросы на выполнение определенных работ в системе и осуществляющие доступ к ресурсам системы с помощью устройств ввода, в качестве которых чаще всего выступают различного рода терминалы;
- внешние управляемые объекты, формирующие с помощью подключенных к ним датчиков и преобразователей информации данные о состоянии объекта, требующие обработки в системе;
- другие системы, рассматриваемые по отношению к исследуемой системе как внешняя среда.

В качестве внутренних источников нагрузки могут выступать:

- отдельные устройства системы;
- программы, выполняемые в ЦП и формирующие запросы ввода-вывода к внешним устройствам;
- ВС, входящие в состав компьютерной сети и формирующие сообщения-ответы, посылаемые другим абонентам сети (ВС и пользователям).

При построении моделей ВС внешние источники нагрузки могут быть включены в состав модели и рассматриваться как внутренние источники и, наоборот, внутренние источники нагрузки могут быть исключены из модели и заменены внешними источниками, что определяется целью исследования, правилом формирования нагрузки и рядом других факторов.

Описание источника нагрузки предполагает задание следующих параметров:

- тип источника (внешний, внутренний);
- тип формируемой в источнике нагрузки;
- интенсивность λ (частота) формирования нагрузки или интервалы времени T между двумя соседними моментами формирования запросов к ресурсу.

Интенсивность λ и средний интервал времени T связаны соотношением: $\lambda = 1/T$.

Нагрузочные параметры. Для описания нагрузки используются следующие основные параметры:

- *интенсивность формирования нагрузки*, определяемая как среднее число запросов, поступающих к ресурсу за единицу времени;
- *ресурсоемкость*, задаваемая в дальнейшем как время обработки запроса (занятия ресурса) τ ;
- *коэффициент нагрузки ресурса* или просто *нагрузка*: $y = \lambda \tau$.

Следует различать нагрузку y и загрузку ρ системы. Загрузка ρ , представляющая собой коэффициент использования системы, никогда не может превысить значения 1, в то время как нагрузка y , характеризующая количество работы, возлагаемой на систему, может быть больше 1.

Вычислительная нагрузка. Вычислительная нагрузка определяет объем работ, выполняемых в ВС, т.е. потребности в ресурсах ВС при решении определенных прикладных задач, и задается совокупностью параметров, определяющих нагрузку во всех устройствах системы.

В зависимости от типа используемых ресурсов вычислительная нагрузка делится на процессорную и периферийную, которые описываются соответствующими параметрами ресурсоемкости, позволяющими определить параметры нагрузки.

Процессорная нагрузка связана с обработкой данных в процессорах системы: ЦП и ПВВ (КВВ).

В процессе системного анализа и проектирования ВС в качестве параметров процессорной обработки могут служить ресурсоемкость (трудоемкость) задачи в процессоре $X_{\text{ЦП}}$, измеряемая числом инструкций или команд, выполняемых в ЦП при решении задачи, и его быстродействие $V_{\text{ЦП}}$ (число команд/инструкций в секунду, MIPS). Тогда время обработки задачи в ЦП: $\tau_{\text{ЦП}} = X_{\text{ЦП}} / V_{\text{ЦП}}$.

Периферийная нагрузка связана с работой внешних (периферийных) устройств: УВВ и ВЗУ.

Ресурсоемкость УВВ определяется объемом данных $G_{\text{УВВ}}$, передаваемых в процессе обмена между УВВ и ОП, и скоростью передачи данных $V_{\text{УВВ}}$ конкретного устройства: $\tau_{\text{УВВ}} = G_{\text{УВВ}} / V_{\text{УВВ}}$.

Ресурсоемкость ВЗУ, являющегося одновременно памятью и устройством, задается следующими параметрами:

- 1) ёмкостью $E_{\text{ВЗУ}}$, занимаемой задачей в ВЗУ;
- 2) временем занятия ВЗУ, которое складывается из:
 - времени доступа $\tau_{\text{д}}$, затрачиваемого на поиск данных на внешнем носителе информации (НМД или НМЛ) и зависящего, в общем случае от расстояния r , на которое перемещается магнитная головка (для НМД) или магнитная лента (для НМЛ), и скорости $V_{\text{п}}$ их перемещения: $\tau_{\text{д}} = r / V_{\text{п}}$;
 - времени поиска $\tau_{\text{н}}$ начала записи, затрачиваемого для НМД на поворот диска (в среднем на пол-оборота) и для НМЛ на перемещение магнитной ленты на величину старт-стопного промежутка;

- времени передачи данных τ_{Π} , определяемого, как и для УВВ, объемом передаваемых данных $G_{ВЗУ}$ и скоростью передачи $V_{ВЗУ}$:
$$\tau_{\Pi} = G_{ВЗУ} / V_{ВЗУ}.$$

Таким образом, время обслуживания запроса в ВЗУ: $\tau_{ВЗУ} = \tau_{д} + \tau_{н} + \tau_{\Pi}$.

Сетевая нагрузка. Сетевая нагрузка создается сообщениями (пакетами) и определяется потребностями в ресурсах при передаче пакетов в компьютерной сети. В качестве ресурсов выступают каналы и узлы связи. Таким образом, сетевая нагрузка складывается из нагрузки, создаваемой пакетами в каналах связи (канальная нагрузка) и в узлах связи (узловая нагрузка).

Канальная нагрузка задается для каждого канала связи (КС) компьютерной сети как $Y_{КС_i} = \lambda_{КС_i} T_{C_i}$, где $\lambda_{КС_i}$ – интенсивность потока пакетов в i -м КС (пакетов в единицу времени); T_{C_i} – среднее время передачи пакета по i -му КС.

Интенсивность потока пакетов в КС определяется суммарной интенсивностью Λ_C передачи пакетов пользователями в сеть и вероятностью $P_{КС_i}$ того, что эти пакеты пройдут через i -й КС: $\lambda_{КС_i} = \Lambda_C P_{КС_i}$.

Время передачи пакета по КС определяется его длиной L_C и пропускной способностью $C_{КС_i}$ соответствующего канала: $T_{КС_i} = L_C / C_{КС_i}$ ($i = \overline{1, n}$), где n – количество КС в сети.

Узловая нагрузка определяет потребность в ресурсах маршрутизатора или коммутатора при обработке пакетов. Обработка заключается в приеме и проверке корректности поступающих пакетов, анализе заголовка и выборе выходного канала связи.

Маршрутизатор и коммутатор можно рассматривать как специализированные вычислительные машины, следовательно, для описания узловой нагрузки могут использоваться те же параметры, что и для вычислительной нагрузки.

4.3. Характеристики вычислительных систем

Эффективность вычислительных систем и компьютерных сетей описывается совокупностью характеристик, которые могут быть разбиты на две группы (рисунок 27):

- качественные;
- количественные.

К качественным характеристикам компьютерных сетей относятся:

- 1) операционные возможности сети, представляющие собой перечень услуг (сервисов) по передаче и обработке данных, предоставляемых пользователям сети, например, передача данных между удаленными пользователями сети, доступ к удаленным файлам, электронная почта и т.д.;

2) масштабируемость – способность сети при ее наращивании (при увеличении ресурсов) линейно увеличивать или, по крайней мере, не уменьшать свою производительность;

3) гибкость – способность системы приспосабливаться к изменяющимся условиям функционирования при выходе из строя оборудования или при изменении нагрузки;

4) совместимость – возможность совместной работы оборудования разных производителей и т.п.

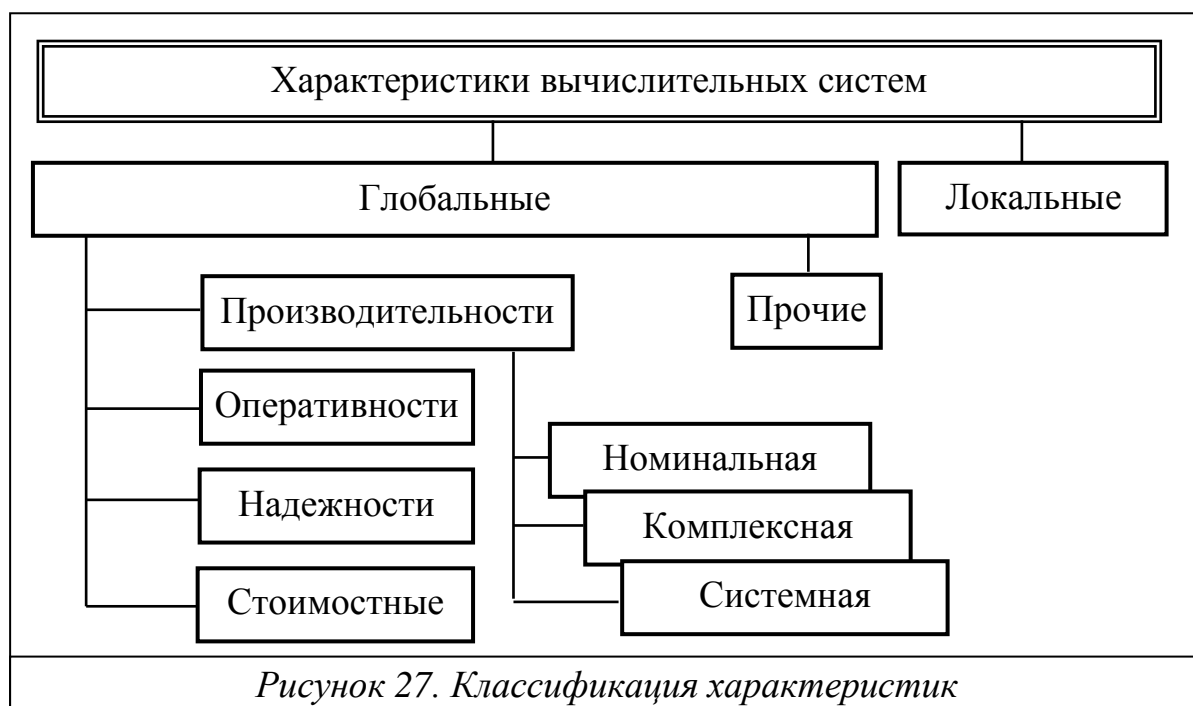


Рисунок 27. Классификация характеристик

Количественные характеристики вычислительных систем и компьютерных сетей представляют собой совокупность показателей эффективности (качества), описывающих процессы обработки и передачи данных.

Количественные характеристики можно разделить на две группы:

- глобальные, определяющие наиболее важные свойства системы как целостного объекта;
- локальные, определяющие свойства отдельных устройств или частей системы и позволяющие получить более детальное представление о качестве функционирования системы.

Часто локальные характеристики входят в состав математических выражений для расчета глобальных характеристик.

К глобальным характеристикам относятся:

- характеристики производительности;
- характеристики оперативности;
- характеристики надежности;
- экономические (стоимостные) характеристики;
- прочие характеристики (энергопотребления, массогабаритные и т.д.).

4.3.1. Характеристики производительности

Производительность ВС – мера мощности, определяющая количество работы, выполняемой системой в единицу времени.

Понятие производительности охватывает широкую номенклатуру показателей эффективности ВС, определяющих качество функционирования как системы в целом, так и отдельных ее подсистем и элементов – технических и программных средств.

Производительность ВС зависит, в первую очередь, от производительности отдельных ее элементов, называемой *скоростью* работы или *быстродействием*. Например быстродействие процессора измеряется числом команд (инструкций, операций), выполняемых в единицу времени.

Для оценки производительности ВС в целом могут использоваться следующие показатели:

- номинальная производительность;
- комплексная производительность;
- системная производительность.

Номинальная производительность характеризуется совокупностью значений быстродействий устройств, входящих в состав ВС: $V=(V_1, \dots, V_n)$, где n – число устройств. Другими словами, номинальная производительность характеризует предельные возможности средств обработки данных, входящих в состав ВС.

При сравнении различных ВС часто используется номинальная производительность, определяемая как суммарное быстродействие процессоров:

$$V_H = \sum_{i=1}^N V_i,$$

где V_i – быстродействие i -го процессора ($i = 1, \dots, N$); N – количество ЦП.

В процессе функционирования ВС возникают конфликты при одновременном обращении ЦП к одним и тем же общим ресурсам: оперативной памяти, внешним устройствам и т.д. При этом реальная производительность ВС оказывается ниже номинальной (предельной) производительности в результате простоев ЦП, ожидающих освобождения занятых общих устройств.

Реальная производительность ВС, характеризующая совместную работу технических средств системы и учитывающая снижение быстродействий отдельных устройств в результате конфликтов при обращении к общим ресурсам, называется *комплексной производительностью ВС*:

$$\hat{V}_K = \sum_{i=1}^N \hat{V}_i,$$

где $\hat{V}_i = \phi_i V_i$ – реальное (фактическое) быстродействие i -го процессора; ϕ_i – коэффициент использования процессора, характеризующий снижение его быстродействия ($0 < \phi_i < 1$).

Наиболее важным показателем производительности ВС, как совокупности технических и программных средств, является *системная*

производительность λ_0 , измеряемая числом задач, выполняемых системой за единицу времени:

$$\lambda_0 = \lim_{T \rightarrow \infty} \frac{m(T)}{T},$$

где $m(T)$ – число задач, выполненных за время T .

Очевидно, что системная производительность зависит от режима функционирования ВС и класса решаемых задач, т.е. вычислительной (рабочей) нагрузки. Ясно также, что системная производительность в значительной степени определяется комплексной производительностью ВС.

4.3.2. Характеристики оперативности

Характеристики оперативности описывают задержки, возникающие при передаче и обработке данных в ВС.

Для оценки оперативности в целом используются следующие показатели:

- время доставки сообщений (пакетов);
- время ответа (реакции, отклика).

Время доставки характеризует эффективность организации передачи данных в компьютерной сети и представляет собой интервал времени, измеряемый от момента ввода сообщения в сеть до момента получения сообщения адресатом-получателем.

Время ответа (реакции, отклика) – интервал времени от момента поступления запроса (сообщения, транзакции) в ВС до момента получения ответа (результата). Время ответа представляет собой *время пребывания запроса* в ВС и характеризует эффективность обслуживания пользователей ВС.

В общем случае, время доставки сообщений, как и время ответа, – величина случайная t , что обусловлено случайным характером процессов поступления и передачи данных в сети. Это время может задаваться *средним значением* T или в *вероятностном виде* как вероятность непревышения некоторого заданного значения T^* : $P(t < T^*)$.

В системах реального времени вместо термина "время ответа" часто используют термин "время реакции".

4.3.3. Характеристики надежности

Надежность – способность ВС сохранять свои наиболее существенные свойства на заданном уровне и выполнять возложенные на нее функции в течение фиксированного промежутка времени при определенных условиях эксплуатации.

В качестве характеристик надежности ВС используются:

- вероятность безотказной работы $P(t)$ – вероятность того, что в течение времени t не произойдет отказа;
- интенсивность отказов λ_0 – среднее число отказов за единицу времени;

- время наработки на отказ T_o – промежуток времени между двумя смежными отказами; величина случайная, а ее среднее значение t_o называется средней наработкой на отказ $t_o = 1/\lambda_o$;

- время восстановления T_b – интервал времени от момента наступления отказа до момента восстановления работоспособности системы; величина случайная и обычно задается средним значением t_b , называемым средним временем восстановления;

- коэффициент готовности K_r – доля времени, в течение которого сеть работоспособна: $K_r = t_o / (t_o + t_b)$.

Величина K_r может трактоваться как вероятность того, что в любой момент времени сеть работоспособна, а $(1 - K_r)$ – вероятность того, что сеть находится в состоянии восстановления.

4.3.4. Экономические характеристики

К экономическим (стоимостным) характеристикам ВС относятся:

- стоимость технических средств S_T – суммарная стоимость устройств, входящих в состав ВС: $S_T = S_1 + \dots + S_n$, где n – количество устройств;

- стоимость информационно-программных средств $S_{и}$ – затраты на приобретение или разработку информационно-программных средств, составляющие в современных ВС значительную долю в общей стоимости сети;

- полная (общая) стоимость системы S_C – суммарная стоимость технических и информационно-программных средств: $S = S_T + S_{и}$;

- стоимость эксплуатации (администрирования) ВС S_A – затраты на поддержку функционирования ВС с требуемым качеством;

- полная стоимость владения (ТСО – Total Cost of Ownership) – суммарные затраты на создание ВС и на ее эксплуатацию: $S = S_C + S_A$;

- стоимость передачи и стоимость обработки данных (решения задачи) определяются объемом и стоимостью используемых ресурсов ВС соответственно при передаче и обработке данных;

4.3.5. Локальные характеристики вычислительных сетей

Локальные характеристики описывают эффективность функционирования отдельных устройств и подсистем ВС:

- ЦП, ВУ и т.д.;
- подсистемы ввода-вывода, ядра ВС (ЦП+ОП) и т.п.;
- узлов и каналов связи.

В качестве локальных характеристик ВС могут использоваться различные показатели эффективности, которые могут быть разбиты на две группы:

- временные, отражающие временные аспекты функционирования системы;
- безразмерные.

К временным характеристикам, в частности, относятся:

- время ожидания освобождения ресурсов ВС;
- время пребывания данных в различных устройствах или подсистемах;
- время доставки (задержки) сообщения при передаче между двумя соседними узлами компьютерной сети и т.п.

К безразмерным характеристикам относятся:

- коэффициенты загрузки устройств ВС и каналов связи;
- число запросов, находящихся в состоянии ожидания освобождения устройства и т.п.

4.4. Базовые модели вычислительных систем

При решении задач системотехнического проектирования вычислительных систем (ВС) и компьютерных сетей в качестве простейших моделей, отображающих процессы обработки и передачи данных и используемых для оценки эффективности функционирования отдельных устройств и подсистем, могут использоваться базовые модели в виде СМО различных классов.

4.4.1. Модель серверной обработки

Рассмотрим простейшую модель серверной обработки.

Положим, что к серверу локальной вычислительной сети (рисунок 28,а) в случайные моменты времени поступают запросы от n пользователей, причем каждый пользователь посылает в среднем m запросов в час независимо от того, сколько запросов уже послано и находится в сервере. Суммарная интенсивность запросов к серверу составляет $\Lambda = n \cdot m$ запросов в час. Производительность сервера, измеряемая средним числом запросов, обрабатываемых за единицу времени, составляет Z запросов в минуту. Запросы, поступившие в систему, хранятся в буферной памяти сервера, емкость которой составляет G мегабайт. Максимальная длина запроса равна g килобайт. Запрос, поступивший к серверу и заставший буферную память заполненной, теряется.

В качестве модели описанной системы может служить СМО с одним устройством, однородным потоком заявок и накопителем ограниченной емкости (рисунок 28,б).

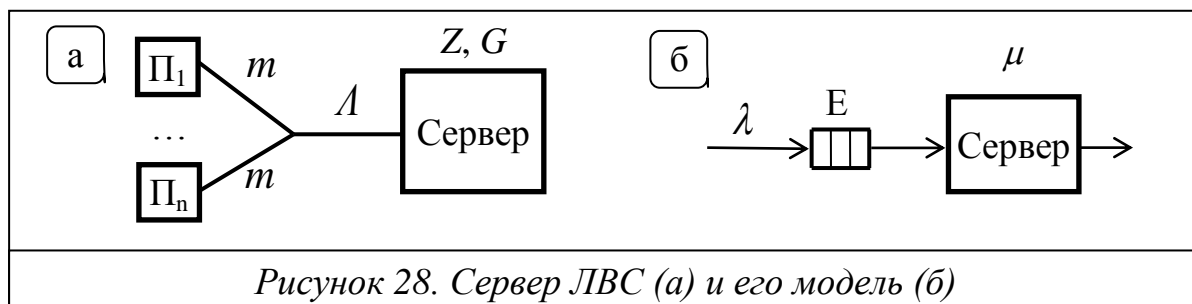


Рисунок 28. Сервер ЛВС (а) и его модель (б)

Выполним параметризацию модели, которая заключается в определении модельных параметров: интенсивности поступления заявок в систему λ ,

интенсивности их обслуживания (обработки) μ и емкости накопителя E . Очевидно, что интенсивность поступления заявок в систему совпадает с суммарной интенсивностью поступления запросов к серверу и с учетом размерностей составляет: $\lambda = \Lambda / 3600 \text{ с}^{-1}$, а интенсивность обслуживания заявок равна производительности сервера: $\mu = Z / 60 \text{ с}^{-1}$. Емкость накопителя может быть рассчитана из условия, что все запросы имеют максимальную длину, по следующей формуле: $E = 1000G/g$. Последнее предположение о максимальной длине запросов позволяет выполнить расчет характеристик функционирования системы, в частности вероятности потерь, для наихудшего случая.

4.4.2. Модель процессорной обработки

Рассмотренная модель серверной обработки является укрупненной и существенно упрощенной, поскольку не учитывает структурную организацию сервера, а, следовательно, не позволяет выявить влияние особенностей реализации процессорной обработки и организации обмена с внешней памятью на характеристики функционирования системы.

При разработке моделей процессорной обработки в ВС предполагается, что в процессе решения задач отсутствует обмен с внешними устройствами системы и, прежде всего, с внешними запоминающими устройствами. Это предположение оправдано в тех случаях, когда ВУ отсутствуют (например, во встроенных системах управления), либо обмен данными с ВУ не значителен и не влияет на качество функционирования системы.

Простейшая модель процессорной обработки отображает функционирование ядра *однопроцессорной* ВС – подсистемы «центральный процессор - оперативная память» (подсистемы ЦП-ОП).

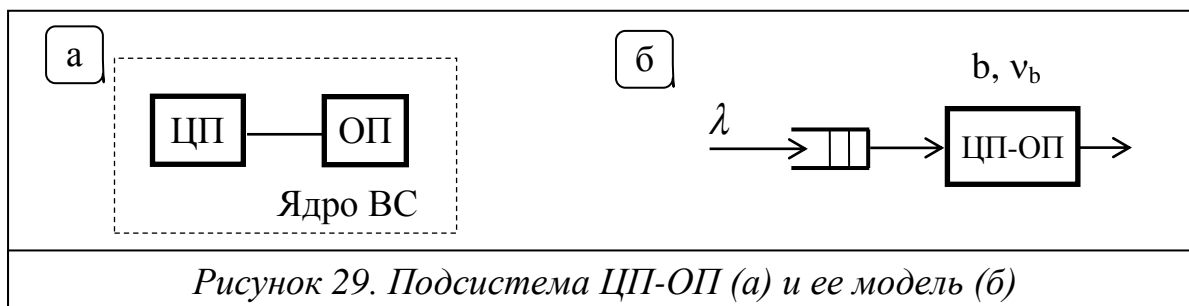
Пусть ВС содержит один ЦП (рисунок 29,а). В системе по запросам пользователей выполняются *однородные задачи*. Запросы на выполнение задач поступают в систему в случайные моменты времени независимо от того, сколько в системе находится запросов, что позволяет рассматривать пользователей как *неограниченный источник запросов*. Средний интервал между запросами равен a . Средняя ресурсоемкость обработки в ЦП одного запроса (ресурсоемкость задач), определяемая как среднее число выполняемых команд, равна θ .

Производительность процессорной обработки (ядра ВС), представляющая собой быстродействие ЦП с учетом быстродействия ОП (времени обращения к ОП) и измеряемая числом команд (инструкций), выполняемых в единицу времени, обозначим через V .

Положим, что размер памяти достаточен для хранения всех поступающих в систему запросов, а точнее, что вероятность переполнения памяти не превышает 10^{-3} . Последнее предположение позволяет использовать в качестве модели процессорной обработки СМО с накопителем неограниченной емкости. При этом отличие (погрешность) результатов расчета характеристик функционирования системы с использованием модели с накопителем

неограниченной емкости от результатов, полученных для модели с накопителем ограниченной емкости, оказывается несущественным и не превышает 5% при условии, что нагрузка системы $y = \lambda\theta/V \leq 0,9$. Если же вероятность переполнения памяти не превышает 10^{-4} , то погрешность результатов менее 5% даже при нагрузке $y = 0,99$.

Таким образом, в качестве модели процессорной обработки при сделанных выше предположениях может использоваться СМО с одним устройством и накопителем неограниченной емкости, в которую поступает однородный поток заявок (рисунок 29,б).



Выполним параметризацию модели и определим следующие модельные параметры:

- интенсивность поступления заявок в систему: $\lambda = 1/a$;
- среднюю длительность их обслуживания (обработки): $b = \theta/V$.

Положим, что запросы, поступающие в систему, образуют простейший поток, а длительность обработки одного запроса распределена по экспоненциальному закону. Тогда в качестве модели процессорной обработки может использоваться СМО типа М/М/1 (в терминах символики Кендалла). Если длительность (ресурсоемкость) обработки запроса распределена по произвольному закону с коэффициентом вариации v_b , то в качестве модели процессорной обработки используется СМО типа М/Г/1.

Базовая модель позволяет рассчитать основные характеристики функционирования подсистемы ЦП-ОП, в частности, загрузку подсистемы, время пребывания запросов (время ответа) и число одновременно находящихся в подсистеме ЦП-ОП запросов.

4.4.3. Модели многомашинной и многопроцессорной обработки

Разработка моделей процессорной обработки в многомашинных и многопроцессорных ВС проводится с учетом основного (с точки зрения функциональной организации) отличия между этими системами, которое заключается в следующем.

При **многомашинной обработке** каждый ЦП может выполнять только те задачи, которые находятся в соответствующем модуле оперативной памяти, доступном только данному процессору (рисунок 30,а). Такие системы называются системами с индивидуальной памятью. Полагая, как и ранее, что в процессе решения задач отсутствует обмен с внешними устройствами, модель процессорной обработки может быть представлена в виде совокупности

одноканальных СМО, число которых равно числу вычислительных машин (ВМ), т.е. числу ЦП (рисунок 30,б).

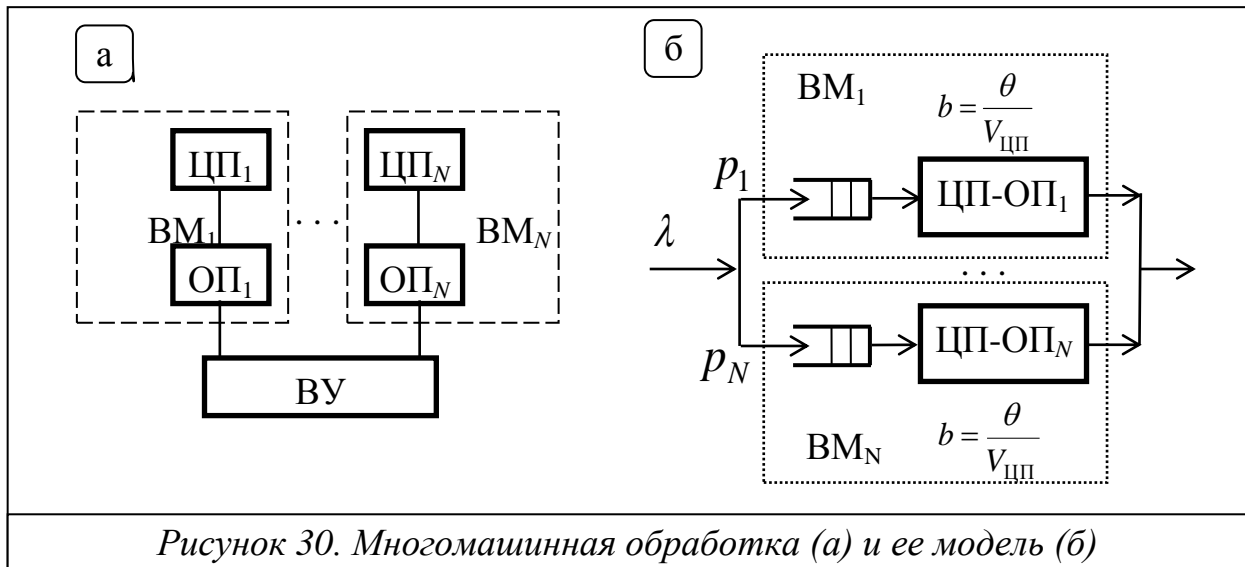


Рисунок 30. Многомашинная обработка (а) и ее модель (б)

При параметризации такой модели кроме интенсивности поступления запросов в систему $\lambda = 1/a$ и средней длительности обработки $b = \theta/V_{\text{ЦП}}$ дополнительно необходимо задать вероятности p_1, \dots, p_N , описывающие долю запросов, направляемых соответственно к ЦП₁, ..., ЦП_N, причем $\sum_{i=1}^N p_i = 1$. Тогда интенсивность потока запросов к ЦП_i рассчитывается как $\lambda_i = p_i \lambda$ ($i = \overline{1, N}$). Если вероятности p_1, \dots, p_N не заданы, то можно воспользоваться предположением о равновероятном обращении к каждому ЦП: $p_i = 1/N$ ($i = \overline{1, N}$). Заметим, что ЦП могут иметь разные производительности, тогда и средние длительности обработки в разных ЦП будут различными: $b_i = \theta/V_{\text{ЦП}_i}$ ($i = \overline{1, N}$).

При **многопроцессорной обработке** все ЦП имеют доступ к любому модулю оперативной памяти и могут выполнять любую задачу, расположенную в общей ОП (рисунок 31,а). Такие системы называются системами с общей оперативной памятью (ООП), а процессорная обработка в них представляется моделями в виде многоканальной СМО, в которой число устройств равно числу ЦП (рисунок 31,б). Обычно предполагается, что все ЦП имеют одинаковую производительность $V_{\text{ЦП}}$, и поступивший в систему запрос может быть обработан любым свободным ЦП.

При расчете длительности обработки запросов в ЦП вместо номинальной производительности $V_{\text{ЦП}}$ необходимо использовать комплексную производительность $\hat{V}_{\text{ЦП}}$, учитывающую конфликты между ЦП при обращении к общей оперативной памяти: $b = \theta/\hat{V}_{\text{ЦП}}$.

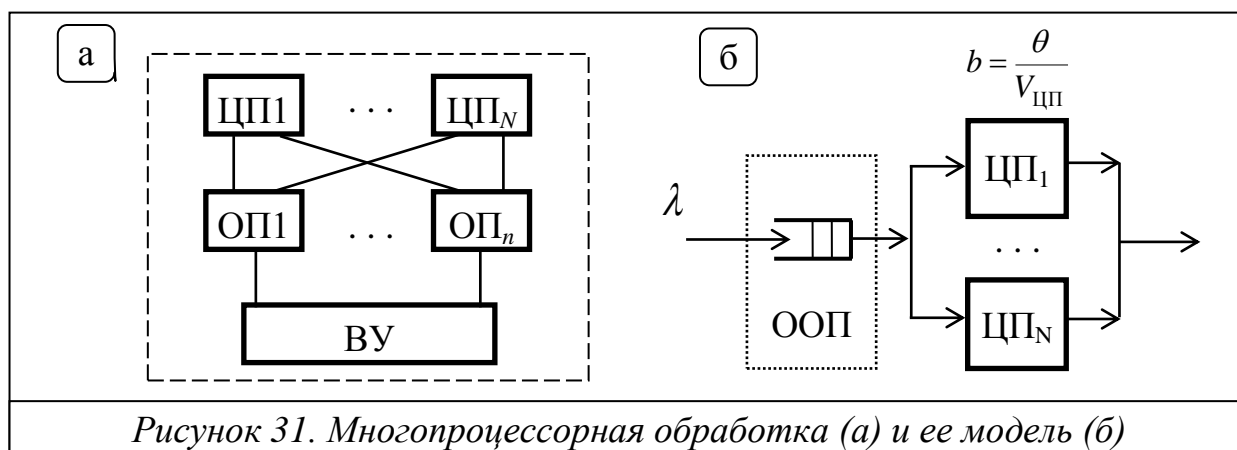


Рисунок 31. Многопроцессорная обработка (а) и ее модель (б)

4.4.4. Модель процессорной обработки с неоднородной нагрузкой

В общем случае в вычислительную систему могут поступать запросы разных типов, образующие неоднородный поток запросов, которые обрабатываются разными прикладными программами $ПП_1, \dots, ПП_N$, формирующими H типов (классов) задач. Эти программы характеризуются, прежде всего, разной ресурсоемкостью: $\theta_1, \dots, \theta_H$ и, возможно, разными требованиями к задержке при их выполнении. Такие вычислительные системы, в отличие от рассмотренных выше, называются системами с неоднородной нагрузкой.

Иногда при построении модели неоднородная нагрузка может быть сведена к однородной путем усреднения параметров, при этом интенсивность потока запросов в систему рассчитывается как сумма интенсивностей потоков

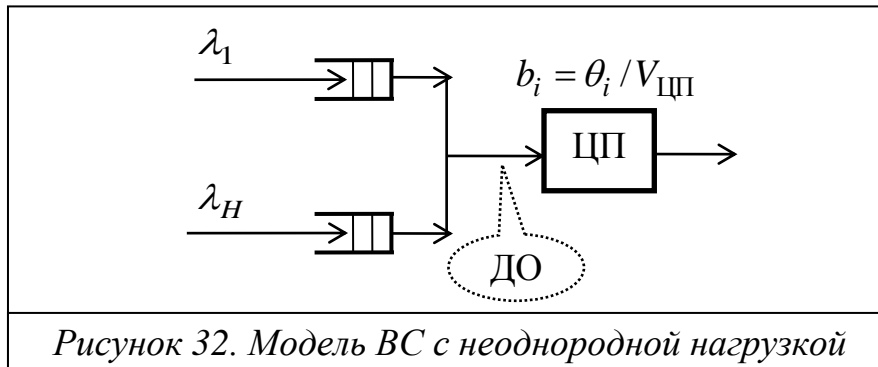
каждого класса: $\Lambda = \sum_{i=1}^H \lambda_i$, где λ_i – интенсивность потока запросов класса

i ($i = \overline{1, H}$), а ресурсоемкость «усредненной» задачи: $\theta = \frac{1}{\Lambda} \sum_{i=1}^H \lambda_i \theta_i$.

В тех случаях, когда в процессе построения модели не удается неоднородную нагрузку свести к однородной, в качестве модели процессорной обработки используются модели в виде СМО с неоднородным потоком запросов на решение задач разных классов (рисунок 32). Невозможность сведения к однородной задаче может быть обусловлена разными причинами, в частности, наличием приоритетов между задачами, наличием разных требований к задержке при выполнении задач, необходимостью дифференцирования характеристик задач разных классов, особенно в случае существенного различия длительностей процессорной обработки в ВС.

В качестве модели процессорной обработки с неоднородной нагрузкой может использоваться СМО с накопителями с ограниченной емкостью или с неограниченной емкостью. Последняя, в частности, может использоваться в случае, если вероятность переполнения памяти не превышает 10^{-4} , а суммарная

нагрузка системы $Y = \frac{1}{\Lambda} \sum_{i=1}^H \frac{\lambda_i \theta_i}{V_{ЦП}} \leq 0,99$.



Модель с неограниченной емкостью накопителей является предпочтительной, поскольку допускает аналитический расчет характеристик функционирования системы при различных дисциплинах обслуживания запросов и произвольных законах распределений длительностей обработки запросов разных классов.

При использовании моделей с неоднородной нагрузкой и накопителями ограниченной емкости аналитический расчет характеристик оказывается затруднительным и, в общем случае, невозможным.

Параметризация модели с неоднородной нагрузкой предполагает задание параметров нагрузки для каждого класса задач ($i = \overline{1, H}$):

- интенсивность поступления запросов на решение задач: $\lambda_i = 1/a_i$;
- средняя длительность обработки запросов: $b_i = \theta_i / V_{\text{ЦП}}$.

Кроме того, необходимо задать дисциплину обслуживания (ДО) запросов разных классов, а в случае модели с накопителями ограниченной емкости – дисциплину буферизации.

Рассматриваемая модель может использоваться для решения задачи функционального проектирования информационно-управляющих систем, заключающегося в определении стратегии управления вычислительным процессом, задаваемой в виде дисциплины обслуживания (ДО) запросов, поступающих в систему.

4.4.5. Модели каналов связи

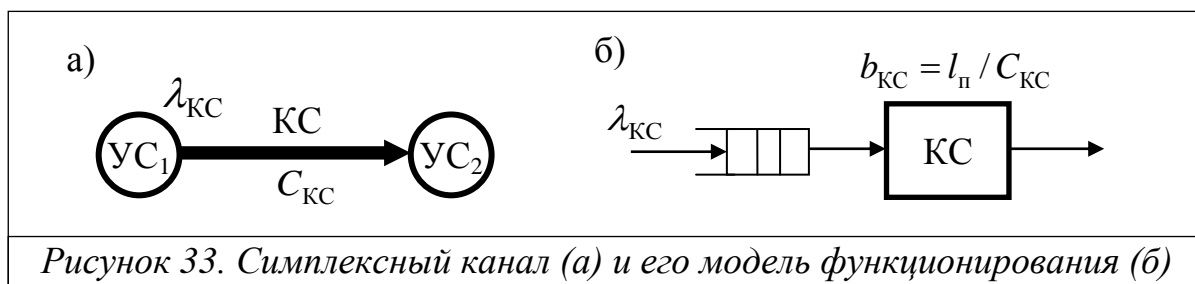
Рассмотрим простейшие модели каналов связи (КС) в компьютерных сетях. В качестве моделей симплексного, полудуплексного и дуплексного КС могут служить СМО.

В соответствии с рекомендациями МСЭ-Т Y.1541, устанавливающими требования к сетевым показателям качества для служб, основанных на протоколе IP, вероятность потери пакетов для любого вида трафика не должна превышать 10^{-3} . Это позволяет в качестве моделей каналов связи использовать СМО с накопителями неограниченной емкости при условии, что нагрузка КС не превышает значения 0,9. В случае высоконагруженных каналов в качестве моделей необходимо использовать СМО с накопителями ограниченной емкости.

Положим, что два узла (компьютеры, маршрутизаторы, коммутаторы) $УС_1$ и $УС_2$ связаны **симплексным каналом** КС (рисунок 33,а), передающим данные (пакеты) только в одном направлении – от $УС_1$ к $УС_2$ с интенсивностью $\lambda_{КС}$. Средняя длина передаваемых пакетов равна $l_{п}$. Пропускная способность КС составляет $C_{КС}$. При этих предположениях модель симплексного канала может быть представлена в виде одноканальной СМО (рисунок 33,б).

Выполним подробную параметризацию модели.

Заявка в СМО соответствует блоку передаваемых данных (пакет, кадр, сообщение) по каналу связи компьютерной сети, а устройство КС отображает задержку, связанную с передачей пакета по КС от $УС_1$ к $УС_2$.



Пакеты, ожидающие освобождения КС, образуют в выходном буфере передающего узла $УС_1$ очередь перед каналом. Если вероятность переполнения буфера меньше 10^{-4} , а нагрузка КС не превышает 0,99, то в качестве модели может использоваться СМО с накопителем неограниченной емкости. При этом погрешность результатов расчета характеристик функционирования не превысит 5%. В противном случае следует воспользоваться моделью с накопителем ограниченной емкости, которая рассчитывается исходя из емкости G выходного буфера передающего узла и максимальной длины l_{max} передаваемых пакетов: $E = G / l_{max}$.

Время $b_{КС}$, в течение которого КС занят в процессе передачи одного пакета, т.е. время обслуживания заявки в модели, складывается из времени распространения сигнала в среде передачи τ_C и времени $\tau_{п}$, затрачиваемого непосредственно на передачу пакета длиной l по КС с пропускной способностью $C_{КС}$:

$$b_{КС} = \tau_C + \tau_{п}.$$

Время распространения сигнала в среде передачи τ_C зависит от расстояния между узлами связи (длины КС) $L_{КС}$ и скорости распространения сигнала, которая для проводных КС в 1,5 – 2 раза меньше скорости света c :

$$\tau_C = \frac{(1,5 \div 2)L_{КС}}{c}$$

Время $\tau_{п}$ передачи пакета по КС зависит от длины пакета $l_{п}$ [бит] и пропускной способности КС $C_{КС}$ [бит/с]:

$$\tau_{п} = \frac{l_{п}}{C_{КС}}.$$

Применяя последнюю формулу, следует учитывать размерности $l_{\text{п}}$ и $C_{\text{КС}}$. Например, если длина пакетов $l_{\text{п}}$ задается в [байтах], а пропускная способность $C_{\text{КС}}$ в [Мбит/с], то время передачи [в секундах] будет рассчитываться по формуле:

$$\tau_{\text{п}} = \frac{8l_{\text{п}}}{10^6 C_{\text{КС}}}.$$

В современных компьютерных сетях с высокоскоростными КС, пропускная способность которых может достигать десятков и более гигабит в секунду, время распространения сигнала оказывается соизмеримым и даже больше, чем время передачи самого пакета.

Действительно, время распространения сигнала на расстоянии $L_{\text{КС}}=1$ км составляет примерно $\tau_{\text{с}}=5$ мкс, а время передачи пакета длиной $l_{\text{п}}=1250$ байт, т.е. 10 тысяч бит, по каналу с пропускной способностью $C_{\text{КС}}=1$ Гбит/с будет равно 10 мкс. Следовательно, при длине КС более 20 км время распространения сигнала превысит на порядок время передачи пакета.

В случае низкоскоростных каналов и при небольших расстояниях между узлами компьютерной сети время $\tau_{\text{п}}$ передачи пакета может значительно превышать время распространения сигнала: $\tau_{\text{п}} \gg \tau_{\text{с}}$. Тогда $b_{\text{КС}} = l_{\text{п}} / C_{\text{КС}}$.

В качестве простейшей модели **полудуплексного канала связи** (рисунок 34,а), передающего данные в обоих направлениях, но в разные интервалы времени, может также использоваться одноканальная СМО, но с неоднородной нагрузкой (рисунок 34,б). Последнее обусловлено тем, что в отличие от симплексного канала интенсивность $\lambda_{\text{КС}}$ передачи пакетов по КС складывается из интенсивностей $\lambda'_{\text{КС}}$ и $\lambda''_{\text{КС}}$ передачи пакетов соответственно в прямом (от узла $У_1$ к узлу $У_2$) и обратном (от узла $У_2$ к узлу $У_1$) направлениях: $\lambda_{\text{КС}} = \lambda'_{\text{КС}} + \lambda''_{\text{КС}}$. Кроме того, необходимо задать метод доступа (МД), в соответствии с которым осуществляется доступ к КС, то есть реализуется последовательность занятия общего канала пакетами, передаваемыми в разных направлениях.

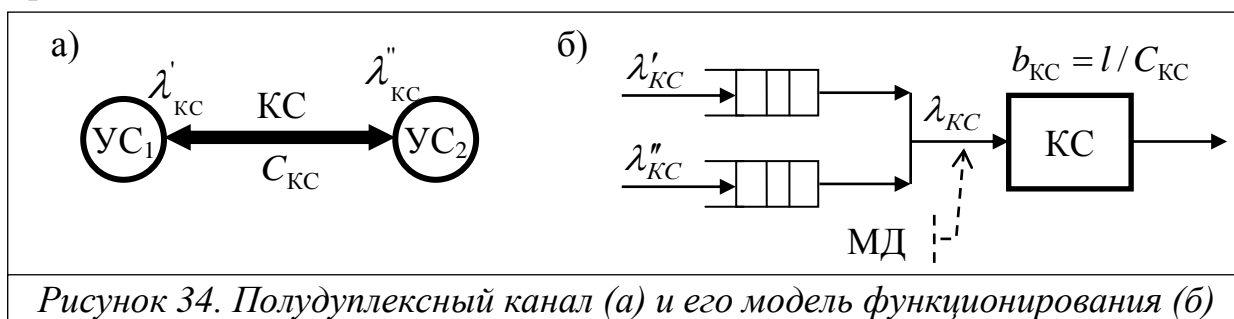


Рисунок 34. Полудуплексный канал (а) и его модель функционирования (б)

Примерами МД могут служить:

- циклические методы, когда в каждом направлении допускается передача группы пакетов заранее установленного определенного размера, например по одному пакету, либо всех пакетов, находящихся в выходном буфере соответствующего узла связи;

- приоритетные методы, когда, например, пакеты в обратном направлении могут передаваться только в случае отсутствия пакетов, подлежащих передаче в прямом направлении и т.д.

Полагая, что пакеты, передаваемые в прямом и в обратном направлениях, имеют одинаковую среднюю длину $l_{\text{п}}$, получим среднее время передачи пакетов (время обслуживания заявки в модели): $b_{\text{КС}} = l_{\text{п}} / C_{\text{КС}}$. Если же пакеты, передаваемые в прямом и обратном направлениях, имеют разную среднюю длину $l'_{\text{п}}$ и $l''_{\text{п}}$ соответственно, то и средние времена передачи пакетов будут различны: $b'_{\text{КС}} = l'_{\text{п}} / C_{\text{КС}}$ и $b''_{\text{КС}} = l''_{\text{п}} / C_{\text{КС}}$.

Модель **дуплексного канала связи**, представляющего собой два симплексных КС, позволяющих передавать пакеты одновременно в обоих направлениях, может быть представлена в виде двух независимых одноканальных СМО (рисунок 35), каждая из которых отображает процесс передачи пакетов по противоположным направлениям.

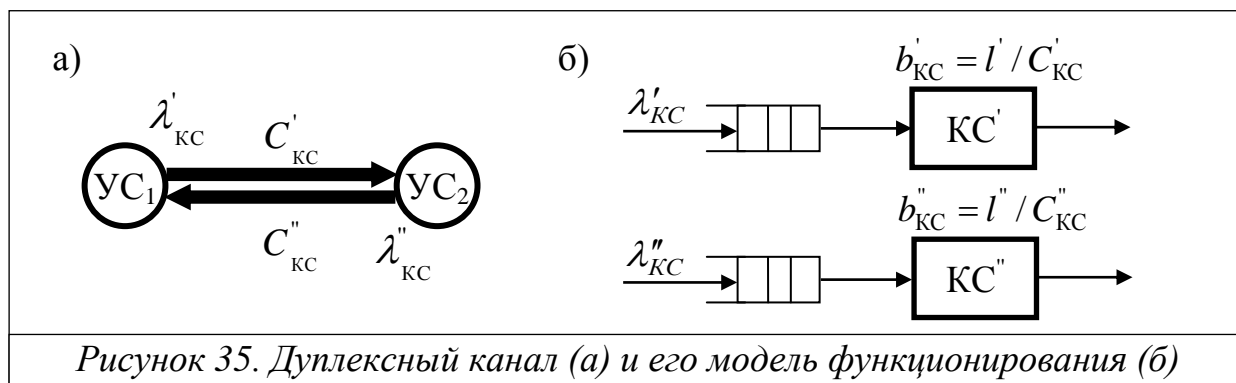


Рисунок 35. Дуплексный канал (а) и его модель функционирования (б)

В общем случае пакеты, передаваемые в прямом (от узла U_1 к узлу U_2) и в обратном (от узла U_2 к узлу U_1) направлениях могут иметь разную среднюю длину $l'_{\text{п}}$ и $l''_{\text{п}}$, а каналы связи $\text{КС}'$ и $\text{КС}''$ – разные пропускные способности $C'_{\text{КС}}$ и $C''_{\text{КС}}$. Соответственно средние времена передачи пакетов по $\text{КС}'$ и $\text{КС}''$ (время обслуживания заявок в модели) будут различны: $b'_{\text{КС}} = l'_{\text{п}} / C'_{\text{КС}}$ и $b''_{\text{КС}} = l''_{\text{п}} / C''_{\text{КС}}$.

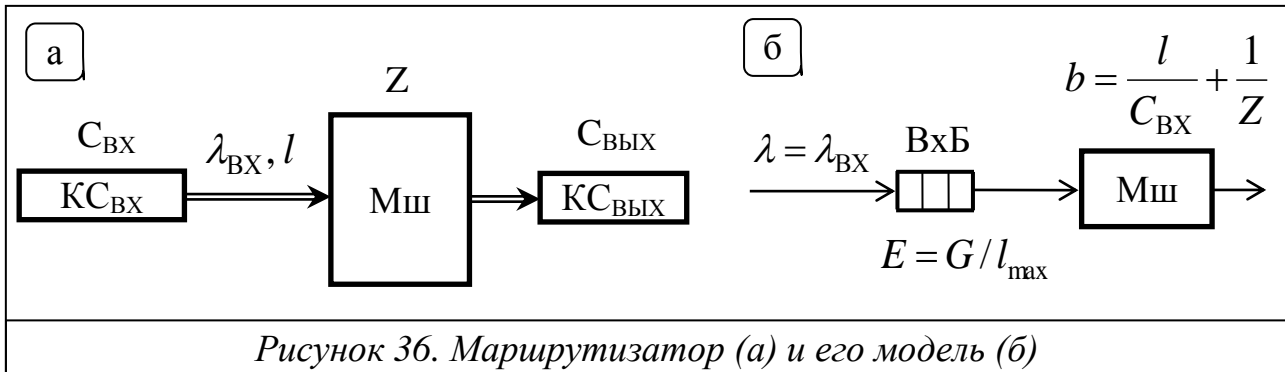
Следует понимать, что рассмотренные модели являются упрощенными и не учитывают многих нюансов, присущих процессам, протекающим в реальных системах. Например, представленная модель дуплексного канала (рисунок 35) не учитывает работу механизма «скользящего окна», реализованного в ряде протоколов канального уровня. В этом случае должны разрабатываться и использоваться более сложные и детальные модели, которые чаще всего не поддаются аналитическому расчету и реализуются обычно средствами имитационного моделирования.

4.4.6. Модели узлов связи

В качестве узлов связи (передачи данных) компьютерных сетей используются маршрутизаторы, коммутаторы 2-го и 3-го уровня и шлюзы, работающие на более высоких уровнях OSI-модели.

Рассмотрим простейшую модель узла связи на примере маршрутизатора, в который по входному каналу $КС_{ВХ}$, пропускная способность которого равна $C_{ВХ}$, с интенсивностью $\lambda_{ВХ}$ поступают пакеты (рисунок 36,а), средняя длина которых равна l . Пакеты поступают во входной буфер маршрутизатора емкостью G , после чего анализируется заголовок пакета и с использованием маршрутной таблицы определяется направление дальнейшей передачи пакета, который и направляется в выходной канал связи $КС_{ВЫХ}$ с пропускной способностью $C_{ВЫХ}$. Производительность маршрутизатора, измеряемая количеством пакетов, обрабатываемых за единицу времени, равна Z .

В качестве простейшей модели маршрутизатора можно воспользоваться одноканальной СМО с накопителем ограниченной емкости (рисунок 36,б).



Выполним подробную параметризацию модели.

Заявки в СМО соответствует поступающим в маршрутизатор пакетам по входному каналу с интенсивностью $\lambda = \lambda_{ВХ}$.

Устройство модели Мш отображает задержку, связанную с обработкой пакета в маршрутизаторе, которая складывается из времени, затрачиваемого на прием (запись) пакета во входной буфер (ВхБ) маршрутизатора τ_B , и времени, затрачиваемого на анализ заголовка пакета и определение направления передачи (выходного буфера) τ_A : $b = \tau_B + \tau_A$.

Время приема пакета, поступающего по $КС_{ВХ}$ во входной буфер, определяется длиной пакета l и скоростью его поступления $C_{ВХ}$: $\tau_B = l/C_{ВХ}$.

Время обработки заголовка пакета определяется производительностью маршрутизатора Z : $\tau_A = \frac{1}{Z}$.

Таким образом, полное время обработки пакета в маршрутизаторе

$$b = \frac{l}{C_{ВХ}} + \frac{1}{Z}.$$

Емкость накопителя в модели приближенно может быть рассчитана исходя из емкости G выходного буфера передающего узла и максимальной длины l_{\max} передаваемых пакетов: $E = G/l_{\max}$. Если вероятность переполнения буфера меньше 10^{-3} или 10^{-4} и нагрузка маршрутизатора $y = \lambda b$ не превышает 0,9 или 0,99 соответственно, то емкость накопителя в модели может считаться неограниченной.

Рассмотренная модель, несмотря на свою простоту, позволяет оценить задержку пакетов в маршрутизаторе. В то же время следует иметь в виду, что эта оценка является приближенной, поскольку не учитывает ряд факторов.

Во-первых, поток пакетов, поступающих по каналу связи, не может быть простейшим, поскольку он обладает последействием, обусловленным тем, что следующий пакет не может поступить по этому же каналу до тех пор, пока во входной буфер не будет принят полностью предыдущий пакет. По этой же причине интенсивность поступления пакетов не может быть больше некоторого значения $\lambda_{\max} = 1/\tau_B$, где $\tau_B = l/C_{\text{ВХ}}$ – среднее время прием пакета во входной буфер маршрутизатора. Так, например, при $l = 1250$ байт и $C_{\text{ВХ}} = 1$ Мбит/с интенсивность $\lambda_{\max} = 100$ пакетов в секунду, а при $C_{\text{ВХ}} = 1$ Гбит/с – $\lambda_{\max} = 100$ тысяч пакетов в секунду.

Во-вторых, процесс приема во входной буфер поступающего пакета и процесс обработки (анализ заголовка пакета и определение направления передачи) ранее принятого пакета могут выполняться параллельно. В этом случае полное время обработки пакета в маршрутизаторе $b < \frac{l}{C_{\text{ВХ}}} + \frac{1}{Z}$. Однако

предположение о том, что $b = \frac{l}{C_{\text{ВХ}}} + \frac{1}{Z}$, позволяет получить верхнюю оценку времени задержки пакетов в маршрутизаторе.

Для того чтобы маршрутизатор работал без перегрузок, необходимо, чтобы его загрузка была меньше единицы: $\rho = \lambda b < 1$.

Положим, что интенсивность поступления пакетов $\lambda = 0,9\lambda_{\max}$. Тогда

$$\rho = 0,9\lambda_{\max} b = 0,9 \frac{C_{\text{ВХ}}}{l} \left(\frac{l}{C_{\text{ВХ}}} + \frac{1}{Z} \right) = 0,9 \left(1 + \frac{C_{\text{ВХ}}}{Zl} \right) < 1,$$

Из последнего неравенства можно получить оценку требуемой производительности маршрутизатора при заданной пропускной способности канала связи и длины передаваемых пакетов:

$$Z > \frac{10C_{\text{ВХ}}}{l}.$$

При $C_{\text{ВХ}} = 1$ Гбит/с и $l = 1250$ байт: $Z > \frac{10 * 10^9}{10^4} = 10^6$, т.е.

производительность маршрутизатора должна составлять не менее 1 млн. пакетов в секунду.

4.5. Сетевые модели вычислительных систем

Для оценки эффективности систем в целом используются сетевые модели, представляемые в виде разомкнутых (РСМО) и замкнутых (ЗСМО) сетей массового обслуживания. Они позволяют учитывать нюансы структурно-функциональной организации и проводить более детальное исследование свойств систем, и, следовательно, более эффективно решать задачи проектирования. Вид сетевой модели зависит от типа источника поступающих в систему запросов. В том случае, когда запросы на обработку (решение задач) или передачу данных поступают из внешнего источника независимо от состояния системы, т.е. от количества запросов уже находящихся в системе, обычно используются РСМО. ЗСМО используются при наличии в исследуемой системе зависимого источника запросов, например, при построении моделей ВС, в которых каждый пользователь посылает в систему новый запрос только после получения ответа на предыдущий запрос.

4.5.1. Модель процессорной обработки в многотерминальной ВС

Рассмотрим многотерминальную вычислительную систему (ВС), работающую в режиме «запрос-ответ», в которой N пользователей с помощью подключенных к системе терминалов посылают запросы, причем каждый пользователь формирует новый запрос только после получения ответа на предыдущий запрос. Положим, что запросы, поступающие в систему, являются однородными и требуют только процессорной обработки, т.е. обмен данными с внешними устройствами отсутствует. Средняя ресурсоемкость процессорной обработки одного запроса равна θ , а производительность (быстродействие) подсистемы ЦП-ОП – V . Пользователь за терминалом, получивший ответ, затрачивает на формирование нового запроса после получения ответа на предыдущий запрос в среднем время τ , которое называется временем обдумывания.

Примерами таких систем могут служить информационно-справочные системы, системы резервирования железнодорожных и авиабилетов, банковские системы и т.п.

В качестве моделей многотерминальных ВС, работающих в режиме «запрос-ответ», применяются замкнутые сетевые модели, в которых пользователи, работающие за терминалами, включаются в состав модели.

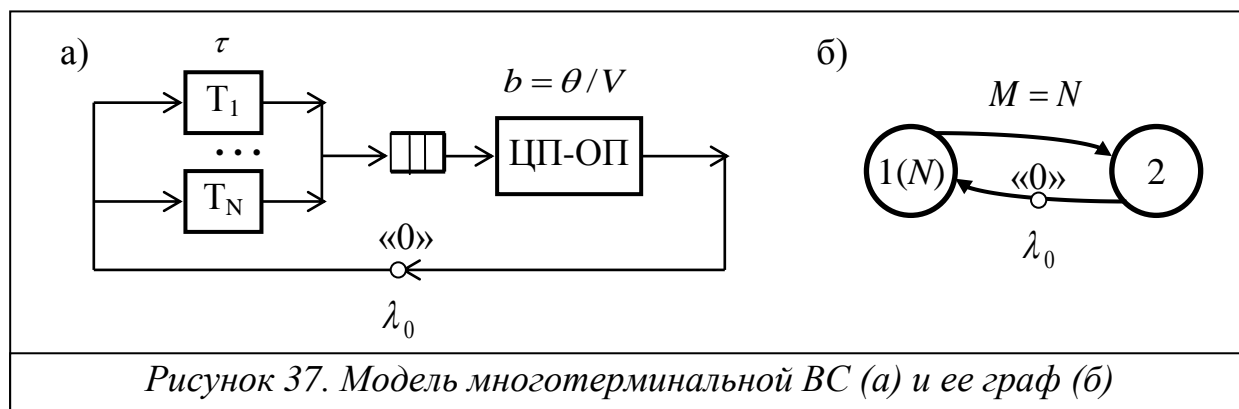
Сетевая модель процессорной обработки в многотерминальной ВС (без учета обмена с внешними устройствами) показана на рисунке 37,а. Модель представляет собой замкнутую двухузловую СМО, граф которой показан на рисунке 37,б.

Заявки, циркулирующие в СМО, соответствуют запросам, формируемым пользователями и поступающим в ВС, и ответам, которые формируются в результате обработки запросов в ЦП-ОП и выдаются пользователям.

Узел 1 отображает работу пользователей за терминалами и представляет собой многоканальную СМО, в которой количество устройств равно количеству активных терминалов N . Среднее время обслуживания в одном устройстве равно среднему времени обдумывания τ , затрачиваемого

пользователем на формирование нового запроса. При этом предполагается, что все пользователи (терминалы) затрачивают в среднем одно и то же время, т.е. все устройства узла являются однородными.

Узел 2 отображает процессорную обработку в подсистеме ЦП-ОП. Средняя длительность обработки запросов $b = \theta/V$.



Выполним параметризацию модели.

Количество заявок, циркулирующих в замкнутой СеМО постоянно и равно количеству *активных* (работающих) терминалов: $M = N$.

Если все пользователи послали в систему по одному запросу, то этой ситуации в модели будет соответствовать состояние, при котором в узле 2 одна заявка находится на обработке в устройстве ЦП-ОП, а остальные $N - 1$ заявок ожидают в накопителе. При этом в узле 1 нет ни одной заявки, т.е. все устройства (терминалы) простаивают.

Если ни один пользователь не послал в систему ни одного запроса, т.е. все пользователи находятся в состоянии обдумывания (формирования нового запроса), то в модели этому будет соответствовать состояние, при котором все N заявок находятся в узле 1, а устройство ЦП-ОП узла 2 простаивает.

Все остальные возможные состояния модели находятся между этими двумя крайними состояниями, когда часть заявок находится в узле 1 (пользователи формируют новые запросы), а остальная часть заявок – в узле 2 (в подсистеме ЦП-ОП), причем их сумма всегда равна N .

Следует обратить внимание на то, что в узле 1 отсутствует накопитель, поскольку число заявок в модели равно числу устройств ($M = N$) в узле 1 и, следовательно, заявка, поступившая в узел 1, всегда найдет свободное устройство. Заметим, что данная модель не отражает того факта, что в реальной системе ответ возвращается именно тому пользователю, который и послал соответствующий запрос. В нашей модели заявка, возвращаясь в узел 1 (к пользователю), может попасть в любое свободное устройство (терминал), что не соответствует действительности. Однако в случае однородных терминалов это несоответствие абсолютно несущественно, поскольку не оказывает влияния на результаты расчета характеристик функционирования системы. Если же время обдумывания разных пользователей различно, то необходимо строить более детальные и, следовательно, более сложные модели, учитывающие поступление ответа именно тому пользователю, который и сформировал соответствующий запрос.

Отметим также, что в узле 2 емкость накопителя может быть ограничена значением $N-1$, поскольку длина очереди не может превысить указанного значения. Очевидно, что если накопитель имеет емкость больше, чем $N-1$, результаты расчета характеристик функционирования системы не изменятся.

Одной из важных характеристик многотерминальных ВС является производительность, измеряемая в реальной системе количеством запросов, обрабатываемых системой за единицу времени. В замкнутой сетевой модели этой характеристике соответствует интенсивность λ_0 движения заявок через нулевую точку «0», выбранную на дуге, выходящей из узла 2 модели в узел 1, которая определяет количество полученных ответов за единицу времени. Относительно нулевой точки также рассчитывается время пребывания заявок в замкнутой модели U , которое складывается из времени обдумывания τ и времени нахождения запроса в подсистеме ЦП-ОП T : $U = \tau + T$. Очевидно, что T представляет собой другую важную для систем типа «запрос-ответ» характеристику – время ответа, которая, следовательно, рассчитывается как $T = U - \tau$.

Рассмотренная замкнутая сетевая модель может быть преобразована в разомкнутую (точнее в простейшую модель процессорной обработки в виде СМО), если пользователи посылают запросы в систему, не ожидая ответа на предыдущие запросы, т.е. не зависимо от того, сколько уже в системе находится запросов. В этом случае терминальный узел 1 можно рассматривать как внешний неограниченный источник заявок, из которого в систему поступают заявки с интенсивностью $\lambda_0 = 1/a$, где a – среднее значение интервала времени между поступающими в систему запросами.

Кроме того разомкнутая модель может быть получена в результате толерантного преобразования замкнутой модели, если загрузка узла 1 много больше загрузки узла 2: $\rho_1 \gg \rho_2$. С учетом того, что $\rho_1 = \frac{\lambda_0 \tau}{N}$ и $\rho_2 = \frac{\lambda_0 \theta}{V}$, последнее неравенство может быть записано в виде: $\tau \gg \frac{N\theta}{V}$, т.е. время обдумывания много больше, чем время обработки N запросов в подсистеме ЦП-ОП. Полагая, что загрузка ρ_1 близка к единице, интенсивность внешнего независимого источника заявок может быть рассчитана как $\lambda_0 \approx \frac{N}{\tau}$. При этом, чем ближе загрузка ρ_1 к единице, тем точнее результаты преобразованной модели.

4.5.2. Модель ВС с внешними устройствами

Современные информационно-вычислительные системы (ИВС) работают с большими по объему базами данных, располагаемыми во внешних запоминающих устройствах (ВЗУ). В процессе функционирования ИВС происходит интенсивный обмен данными с НМД, что оказывает существенное влияние на характеристики функционирования системы. При исследовании и проектировании таких систем необходимо применять модели, отображающие

не только процессорную обработку, но и обмен с внешними устройствами. Процесс функционирования таких систем, связанный с обработкой запросов или решением задач, можно представить как последовательное чередование этапов счета в ЦП и этапов обмена данными с ВЗУ.

Положим, что однородные запросы на решение задач поступают в ИВС из внешнего независимого источника (т.е. независимо от того, сколько уже находится в системе запросов) с интенсивностью $\lambda_0 = 1/a$, где a – среднее значение интервала времени между поступающими в систему запросами. Предположение о независимости источника запросов позволяет использовать в качестве модели ИВС разомкнутую СеМО.

В состав ИВС входят один ЦП с ОП и K внешних запоминающих устройств, в качестве которых будем рассматривать накопители на магнитных дисках (НМД). Положим, что все НМД могут выполнять обмен данными *параллельно и независимо* друг от друга.

Производительность подсистемы ЦП-ОП равна V .

Каждый НМД i ($i = \overline{1, K}$) описывается следующей совокупностью параметров:

- время доступа $\tau_{Д_i}$, затрачиваемое на позиционирование головок чтения-записи (перемещение на требуемую дорожку);
- скорость вращения диска $Q_{МД_i}$;
- скорость передачи данных в процессе обмена $V_{МД_i}$;
- емкость НМД $G_{МД_i}$.

Ресурсоемкость процессорной обработки одного запроса составляет θ .

Положим, что в процессе обработки запроса выполняется D_i обращений к файлам (наборам данных), размещенных в НМД i ($i = \overline{1, K}$), и при каждом обращении к ним передается блок данных, средний размер которого составляет g_i . Очевидно, что изменение значений D_i ($i = \overline{1, K}$) может быть реализовано за счет перераспределения файлов по накопителям внешней памяти.

Модель функционирования ИВС с учетом обмена данными с НМД в виде разомкнутой СеМО показана на рисунке 38,а), а ее граф – на рисунке 38,б).

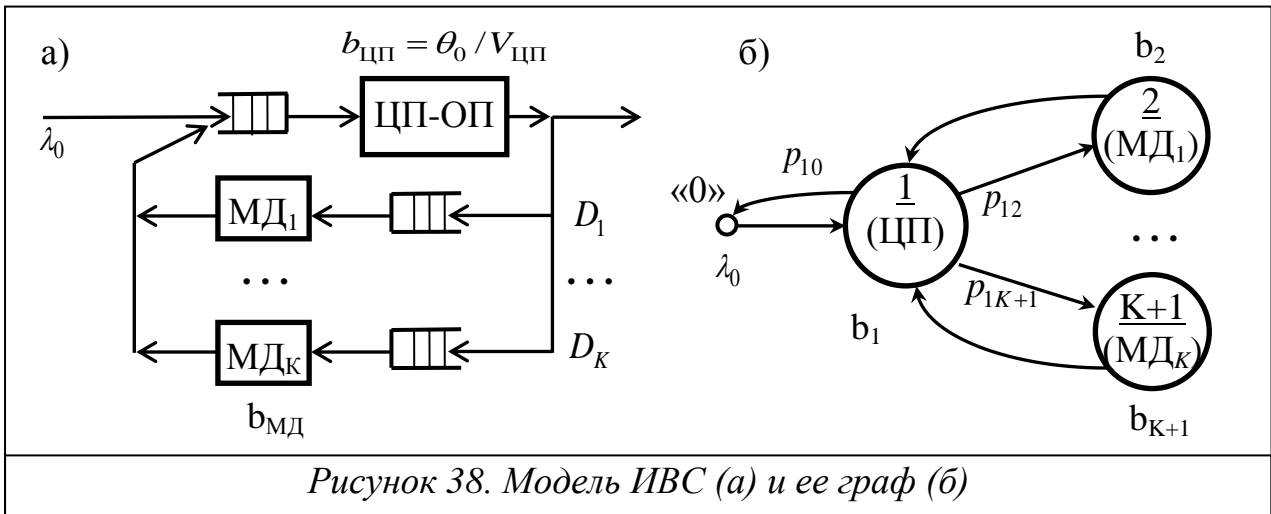
Внешняя среда, откуда заявки поступают в СеМО и куда они возвращаются после обслуживания в сети, на графе модели (рисунок 38,б) обозначена как дополнительный нулевой узел «0».

Узел 1 (ЦП) отображает обработку запросов в подсистеме ЦП-ОП, а узлы 2 (МД $_1$), ..., $K+1$ (МД $_K$) отображают обмен данными соответственно с НМД $_1$, ..., НМД $_K$. Обозначим средние длительности обслуживания заявок в узлах сетевой модели, значения которых должны быть определены в процессе параметризации, как b_1 для узла 1 (ЦП) и b_2, \dots, b_{K+1} для узлов 2 (МД $_1$), ..., $K+1$ (МД $_K$) соответственно.

Выполним параметризацию модели.

Заявки в модели соответствуют поступающим в систему запросам и образуют однородный поток с заданной интенсивностью $\lambda_0 = 1/a$. Заявки в

модели перемещаются в соответствии с последовательностью выполнения обработки запросов в разных устройствах ИВС.



Процесс обработки начинается и заканчивается в ЦП. В отличие от простейших моделей процессорной обработки, в которых не учитывается обмен данными с внешними устройствами (НМД), в сетевой модели процессорная обработка отображается как многоэтапная. Этап процессорной обработки между двумя обращениями к НМД представляет собой непрерывный этап счета в ЦП. Количество этапов счета равно $(D+1)$, где $D = \sum_{i=1}^K D_i$ – количество обращений к НМД в процессе обработки одного запроса. Соответственно, ресурсоемкость этапа счета: $\theta_0 = \theta / (D+1)$ и время обслуживания заявки в узле модели 1 (ЦП) равно $b_1 = b_{\text{ЦП}} = \frac{\theta_0}{V_{\text{ЦП}}} = \frac{\theta}{(D+1)V_{\text{ЦП}}}$

При поступлении запроса к НМД заявка в модели попадает в соответствующий узел, отображающий задержку при обмене данными с НМД.

Длительность обслуживания запросов в НМД складывается из времени, затрачиваемого на выполнение подготовительных операций (поиск данных) τ' , и времени, затрачиваемого непосредственно на передачу данных τ'' . Время τ' складывается из времени доступа τ_d и времени, затрачиваемого на подвод начала считываемого блока данных под головку чтения-записи, которое в среднем равно времени полуоборота магнитного диска: $\tau' = \tau_d + \frac{0,5}{Q_{\text{МД}}}$. Время передачи данных τ'' зависит от длины g передаваемого блока данных и скорости передачи: $\tau'' = \frac{g}{V_{\text{МД}}}$. Таким образом, длительность обслуживания

заявок в узлах $j = 2, \dots, K + 1$ модели рассчитывается по формуле:

$$b_j = \tau_{Д_j} + \frac{0,5}{Q_{МД_j}} + \frac{g_j}{V_{МД_j}}.$$

При описании сетевых моделей в качестве одного из параметров, определяющих структуру модели, используется матрица вероятностей передач $\mathbf{P} = [p_{ij} \mid i, j = 0, 1, \dots, n]$, где p_{ij} – вероятность передачи заявки из узла i в узел j .

Для рассматриваемой модели ИВС (рисунок 38) с заданной конфигурацией связей между узлами матрица вероятностей передач примет вид:

$$\mathbf{P} = \begin{array}{c|cccccc} & 0 & 1 & 2 & \dots & n \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & p_{10} & 0 & p_{12} & \dots & p_{1n} \\ 2 & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ n & 0 & 1 & 0 & \dots & 0 \end{array},$$

причем $\sum_{j=0}^n p_{ij} = 1$ ($i = \overline{0, n}$), поскольку предполагается, что заявки в модели, а, следовательно, и запросы в ИВС, не размножаются и не теряются, т.е. сетевая модель является линейной.

Вероятности p_{1j} ($j = 0, 2, \dots, n$) зависят от количества обращений к НМД в процессе обработки одного запроса и рассчитываются по формулам:

$$p_{10} = \frac{1}{D+1} \text{ и } p_{1j} = \frac{D_j}{D+1} \quad (j = 2, \dots, n).$$

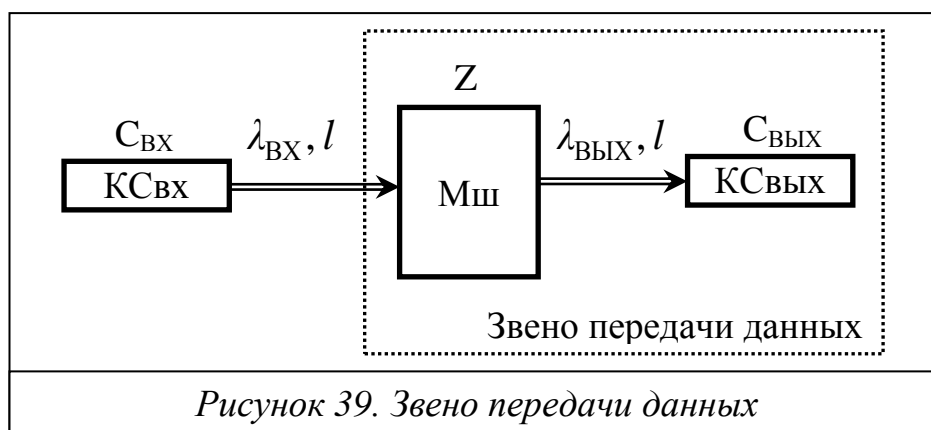
На основе вероятностей передач определяются значения коэффициентов передач α_j ($j = \overline{1, K+1}$), которые используются при расчете сетевых характеристик по формулам (20) – (23).

Для рассматриваемой модели значения коэффициентов передач будут равны: $\alpha_1 = D+1$ и $\alpha_j = D_j$ для $j = 2, \dots, n$. Другими словами, коэффициенты передач для каждого узла, отображающего обмен данными с НМД, будут равны количеству обращений к соответствующему НМД в процессе обработки одного запроса, а коэффициент передачи для узла 1(ЦП), отображающего обработку запроса в подсистеме ЦП-ОП, – количеству этапов счета. Этот факт становится очевидным, если вспомнить, что коэффициент передачи характеризует среднее число попаданий заявки в соответствующий узел за время нахождения в сетевой модели. Таким образом, для рассматриваемой сетевой модели нет необходимости в определении значений вероятностей передач.

Отметим, что если при наличии в ИВС нескольких ВУ невозможен параллельный обмен данными, то в модели будет изображаться только один узел ($K=1$).

4.5.3. Модель звена передачи данных

Рассмотрим звено передачи данных, включающее узел связи (маршрутизатор), в который поступают пакеты по входному каналу связи (КС) и после обработки направляются в соответствующий выходной канал для передачи в смежный узел (рисунок 39). Модель звена передачи данных отображает обработку пакетов в маршрутизаторе и их передачу по выходному каналу.



Пропускные способности входного $КС_{ВХ}$ и выходного $КС_{ВЫХ}$ каналов соответственно равны $C_{ВХ}$ и $C_{ВЫХ}$, а пропускная способность маршрутизатора, измеряемая количеством пакетов, которые маршрутизатор обрабатывает за единицу времени, равна Z .

Положим, что известна интенсивность поступления пакетов $\lambda_{ВХ}$ в маршрутизатор по входному каналу $КС_{ВХ}$. Средняя длина пакетов, передаваемых по КС равна l .

Выше были рассмотрены простейшие модели канала связи (КС) и узла связи (маршрутизатора). Модель звена передачи данных объединяет обе эти модели и представляется в виде разомкнутой двухузловой СеМО (рисунок 40,а), граф которой представлен на рисунке 40,б).

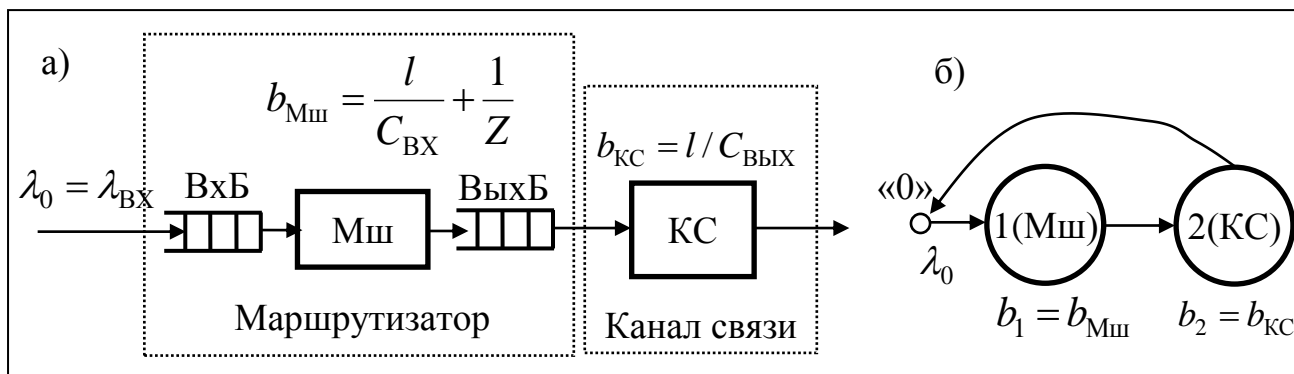


Рисунок 40. Модель звена передачи данных (а) и ее граф (б)

Заявки (пакеты), поступающие в СеМО с интенсивностью $\lambda_0 = \lambda_{\text{ВХ}}$, попадают в очередь, отображающую входной буфер (ВхБ) маршрутизатора, где ожидают освобождения процессора маршрутизатора, занятого обработкой ранее поступившего пакета. Среднее время обработки пакета складывается из времени приема пакета во входной буфер и времени, затрачиваемого на анализ заголовка пакета и определение направления передачи: $b_{\text{Мш}} = \frac{l}{C_{\text{ВХ}}} + \frac{1}{Z}$. Затем

пакет перемещается в выходной буфер (ВыхБ) маршрутизатора, где ожидает освобождения канала связи КС, если он занят передачей ранее обработанного пакета. Пакеты, ожидающие освобождения КС, образуют очередь перед каналом, формируемую в выходном буфере маршрутизатора. Процесс передачи пакета по каналу связи отображается в модели в виде задержки заявки в узле КС на время, учитывающее время распространения сигнала в среде передачи и время непосредственно передачи пакета: $b_{\text{КС}} = \frac{(1,5 \div 2)L_{\text{КС}}}{c} + \frac{l}{C_{\text{ВЫХ}}}$, где $L_{\text{КС}}$ –

длина канала связи, а c – скорость света. Завершение передачи пакета по КС отображается в виде заявки, покидающей СеМО и возвращающейся во внешнюю среду, отображаемую в виде узла «0».

4.5.4. Модель сети передачи данных

Модель сети передачи данных (СПД) представляет собой совокупность взаимосвязанных моделей каналов и узлов связи, отображающую состав и топологию СПД, и изображается в виде СеМО. Размерность такой модели зависит от размера исследуемой СПД и уровня детализации, используемого для представления процессов, протекающих в СПД.

Пакеты, передаваемые в СПД, представляются в сетевой модели заявками, перемещаемыми между узлами сетевой модели в соответствии с логикой передачи пакетов в реальной СПД. Поскольку интенсивность поступления пакетов в СПД не зависит от того, сколько пакетов уже находится в сети, то количество пакетов, находящихся одновременно в сети, может быть любым, что позволяет представить СПД в виде разомкнутой СеМО.

Параметризация модели СПД проводится аналогично параметризации рассмотренных выше моделей каналов и узлов связи, но в виду большой размерности модели отличается громоздкостью и сопровождается большим объемом вычислений.

5. Вопросы для самопроверки

- Какие принципы лежат в основе проектирования систем на системном уровне?
- Что является исходными данными для системотехнического проектирования систем?
- Что определяется в процессе системотехнического проектирования систем?
- Какие уровни моделей в соответствии с принципом иерархического многоуровневого моделирования можно выделить в зависимости от структурно-функциональных особенностей системы?
- Какие величины относятся к параметрам?
- Способ реализации возложенных на систему функций называется ...
- Какие параметры используются для описания нагрузки?
- Что относится к глобальным характеристикам системы?
- Что представляет собой максимальная или предельная производительность системы?
- Что может использоваться в качестве характеристик надежности системы?
- Как называется характеристика, учитывающая как затраты на создание системы, так и затраты на ее эксплуатацию?
- Какие значения может принимать нагрузка системы?
- Какие значения может принимать нагрузка?
- Решение каких основных задач предполагает процесс проектирования ВСС?
- Как называется модель, предназначенная для выявления наиболее существенных аспектов структурно-функциональной организации системы, учет которых необходим для получения требуемых результатов?
- Применение каких методов предполагает математическое моделирование систем?
- Проверка адекватности модели исследуемой системе называется ...
- Что понимается под верификацией модели?
- Синонимом термина "корректировка модели" является ...
- Планирование, состоящее в выборе определенных сочетаний параметров и последовательности проведения экспериментов с использованием методов теории планирования экспериментов, называется ...
- Планирование, направленное на уменьшение времени выполнения одного эксперимента при обеспечении статистической достоверности результатов имитационного моделирования, называется ...
- Какие требования предъявляются к модели?
- Установление соответствия между значениями системных и модельных параметров и характеристик называется ...
- Какие параметры СМО относятся к нагрузочным?

- Какая характеристика показывает степень использования устройства ВС?
- Как называются величины, описывающие эффективность системы?
- Как называется принцип, заключающийся в возможности отображения многих различных систем с помощью одной и той же модели и в возможности представления одной и той же системы множеством различных моделей в зависимости от целей исследования?
 - Какие способы применяются для описания структуры системы?
 - Замещение одного исходного объекта другим объектом и проведение с ним экспериментов с целью получения информации об исходном объекте называется ...
 - Какие модели допускают количественное исследование свойств систем и процессов?
 - Моделирование предоставляет возможность исследования объектов, прямой эксперимент с которыми ...
 - Как называется система с большим числом входящих в его состав элементов и связей между ними?
 - Какие способы используются для описания структуры системы?
 - Как называется правило достижения поставленной цели, описывающее поведение системы и направленное на получение результатов, предписанных назначением системы?
 - Как называется способ описания функции системы в виде последовательностей шагов, которые должна выполнять система для достижения поставленной цели?
 - Как называется способ описания функции системы в виде математических зависимостей в терминах некоторого математического аппарата?
 - Процесс определения свойств, присущих системе называется...
 - К характеристикам системы относятся величины, описывающие её...
 - Способы описания функции системы.
 - Величины, описывающие первичные свойства системы и являющиеся исходными данными при решении задач анализа называются ...
 - Какой метод моделирования является универсальным?
 - Способ достижения поставленной цели за счет выбора определенной структуры и функции системы называется ...?
 - Как называется свойство системы, заключающееся в том, что она рассматривается как единое целое, состоящее из взаимодействующих элементов, возможно неоднородных, но одновременно совместимых?
 - Какие утверждения являются неверными?
 - Величины, описывающие вторичные свойства системы и определяемые в процессе решения задач анализа называются ...
 - Какие свойства присущи сложной системе?
 - К параметрам системы относятся величины, описывающие ...

- Какие величины относятся к внутренним параметрам?
- Какие величины относятся к внешним параметрам?
- Какие величины являются глобальными характеристиками технических систем?
 - Какой метод позволяет выполнять исследование систем на моделях любой степени детализации?
 - Наличие качеств, присущих системе в целом, но не свойственных ни одному из ее элементов в отдельности называется ...
 - Мера одного свойства системы называется...?
 - Степень соответствия системы своему назначению называется ...
 - Процесс порождения функций и структур, удовлетворяющих требованиям, предъявляемым к эффективности системы называется ...
 - Процесс определения свойств, присущих системе, называется ...
 - Мера эффективности системы, обобщающая все свойства системы в одной оценке, называется ...
 - Как называется критерий эффективности, значение которого возрастает при увеличении эффективности системы ?
 - Как называется критерий эффективности, значение которого уменьшается при увеличении эффективности системы?
 - Как называется система, которой соответствует минимальное значение инверсного критерия эффективности?
 - Как называется система, которой соответствует максимальное значение прямого критерия эффективности?
 - Процесс, протекающий в системе - это ...
 - Как называется причина, вызывающая переход процесса из состояния в состояние?
 - Как называются процессы, для которых характерен плавный переход из состояния в состояние?
 - Как называются системы, для которых характерен плавный переход из состояния в состояние?
 - Как называются процессы, для которых характерен скачкообразный переход из состояния в состояние?
 - Как называются процессы, для которых характерен скачкообразный переход из состояния в состояние?
 - Как называется процесс, поведение которого может быть предсказано заранее?
 - Как называется процесс, поведение которого невозможно предсказать заранее?
 - Как называется режим функционирования системы, при котором характеристики системы не зависят от времени?
 - Как называется режим функционирования системы, при котором характеристики системы зависят от времени?
 - С чем связан переходной режим функционирования системы?

• Чем может быть обусловлен неустановившийся режим функционирования системы?

• Как называется режим функционирования, при котором система не справляется с возложенной на нее нагрузкой?

• Какие требования предъявляются к модели?

• Соответствие модели оригиналу, характеризующее степень близости свойств модели свойствам исследуемой системы, называется ...?

• От чего зависит адекватность математических моделей?

• Моделирование может проводиться в условиях неопределенности, обусловленных:

• Что является синонимом понятия "вероятностная модель"?

• Что является антонимом понятия "детерминированная модель"?

• Что является синонимом понятия "содержательная модель"?

• Что является синонимом понятия "концептуальная модель"?

• Что является синонимом понятия "математическая модель"?

• Какие модели являются абстрактными?

• Что является синонимом понятия "материальная модель"?

• Как называется модель, представляющая собой словесное описание наиболее существенных особенностей структурно-функциональной организации исследуемой системы?

• Как называется модель, эквивалентная или подобная оригиналу или процесс функционирования которой такой же, как у оригинала и имеет ту же или другую физическую природу?

• Недостатки статистического моделирования.

• Основное достоинство статистического моделирования.

• Какие методы математического моделирования получили наиболее широкое применение при исследовании технических систем с дискретным характером функционирования?

• Какие элементы входят в состав базовой модели?

• Что представляет собой процесс обслуживания заявки в устройстве СМО?

• В каких случаях оправдано предположение о неограниченной ёмкости накопителя в СМО?

• Правило выбора заявок из очереди для обслуживания в устройстве - это ...?

• Что является основной характеристикой потока заявок?

• Среднее число заявок, проходящих через некоторую границу за единицу времени, называется ...?

• Что представляет собой величина, обратная среднему интервалу между последовательными заявкам в потоке?

• Что представляет собой величина, обратная интенсивности потока заявок?

- В какой последовательности выполняются этапы проектирования системы?
- Укажите последовательность решения задач в процессе исследования сложных систем.
- Какой поток заявок называется однородным? В каких случаях поток заявок в СМО является неоднородным?
- В каких случаях заявки в СМО относятся к разным классам?
- Нарисовать одноканальную СМО с неоднородным потоком заявок. Какие параметры необходимо задать для её описания? Какие характеристики функционирования СМО могут быть рассчитаны по этим параметрам?
- Нарисовать многоканальную СМО с неоднородным потоком заявок. Какие параметры необходимо задать для её описания? Какие характеристики функционирования СМО могут быть рассчитаны по этим параметрам?
- Как называется стационарный ординарный поток без последствия?
- Когда поток заявок является стационарным? Привести примеры нестационарного потока заявок.
- Какой поток заявок называется ординарным? Привести примеры неординарного потока заявок.
- Каким является поток, в котором момент поступления очередной заявки не зависит от того, когда и сколько заявок поступило до этого момента?
- В чём проявляется наличие последствия в потоке заявок? Привести примеры потоков заявок с последствием.
- Понятие интенсивности потока и ее размерность. Что характеризует величина обратная интенсивности?
- По какому закону распределены интервалы времени между заявками в простейшем потоке?
- Какими замечательными особенностями обладает простейший поток заявок?
- Чему равны математическое ожидание, коэффициент вариации и дисперсия интервалов времени между соседними заявками в простейшем потоке, интенсивность которого равна 5 заявок в секунду?
- В систему поступают заявки с интервалом 100 секунд. Чему равно среднее число заявок, которые поступят в систему в течение одного часа в случае: а) детерминированного потока; б) простейшего потока; в) случайного потока?
- В систему поступают заявки двух классов со средним интервалом между соседними заявками 0,5 с и 5 с соответственно. Определить суммарную интенсивность поступления заявок в систему. По какому закону распределены интервалы между заявками суммарного потока?
- В систему поступают заявки трех классов со средним интервалом между соседними заявками 0,1 с; 0,2 с и 0,5 с соответственно. Определить суммарную интенсивность поступления заявок в систему. Чему равен коэффициент вариации интервалов между заявками суммарного потока?

• В двухканальную СМО поступает простейший поток заявок с интенсивностью 15 заявок в секунду, причем с вероятностью $1/3$ заявка направляется ко второму устройству. Чему равна интенсивность потока заявок к первому устройству? Чему равен коэффициент вариации интервалов между заявками потока к первому устройству?

• Что понимается под обслуживанием заявок в СМО? Что такое интенсивность обслуживания заявок в СМО, и какова её размерность?

• Чему равны математическое ожидание, коэффициент вариации и дисперсия длительности обслуживания заявок в СМО, распределенной по экспоненциальному закону, если известно, что интенсивность обслуживания равна 4 заявки в секунду?

• Перечислить возможные дисциплины буферизации. В каких СМО не используются дисциплины буферизации?

• Какие дисциплины обслуживания заявок относятся к беспriorитетным?

• Краткая характеристика приоритетных дисциплин обслуживания заявок.

• Что такое динамические приоритеты?

• Что характеризуют нагрузка и загрузка? В чём отличие загрузки от нагрузки? В каких случаях нагрузка совпадает с загрузкой?

• Перечислить факторы, обуславливающие нестационарный режим работы СМО.

• Что такое и чем характеризуется перегрузка системы? При каких условиях возникают перегрузки системы? В каких СМО не возникают перегрузки?

• При каком условии в одноканальной СМО отсутствуют перегрузки?

• Перечислить характеристики одноканальной и многоканальной СМО с однородным потоком заявок и записать соотношения, устанавливающие их взаимосвязь.

• Перечислить характеристики одноканальной СМО с неоднородным потоком заявок и записать соотношения, устанавливающие их взаимосвязь.

• Почему в СМО, работающей в стационарном режиме, могут возникать очереди? В каких случаях они не возникают? Перечислите причины, обуславливающие возникновение очередей в СМО, работающей в стационарном режиме.

• Какая СеМО называется линейной? Перечислить факторы, обуславливающие нелинейность СеМО.

• Основные отличия замкнутых СеМО от разомкнутых.

• Какая СеМО называется экспоненциальной? Перечислить факторы, обуславливающие неэкспоненциальность СеМО.

• Какая СеМО называется неоднородной? Перечислить факторы, обуславливающие неоднородность СеМО.

• Перечислить параметры разомкнутой и замкнутой однородной неэкспоненциальной СеМО.

- Перечислить параметры разомкнутой и замкнутой неоднородной приоритетной СеМО.
- Каким условиям должны удовлетворять элементы матрицы вероятностей передач в СеМО?
 - Узловые характеристики однородных СеМО и их взаимосвязь.
 - Сетевые характеристики разомкнутых и замкнутых однородных СеМО и их взаимосвязь.
 - Что такое "производительность замкнутой СеМО"? Какие соотношения используются для расчета производительности замкнутой СеМО?

Список литературы

1. Алиев Т.И. Основы моделирования дискретных систем. Учебное пособие. – СПб.: СПбГУ ИТМО, 2009. – 363 с. ISBN 978-5-7577-0336-7.
2. Жожикашвили В.А., Вишнеvский В.М. Сети массового обслуживания. Теория и применение к сетям ЭВМ. - М.: Радио и связь, 1988. - 192 с.: ил.
3. Клейнрок Л. Вычислительные системы с очередями. - М.: Мир, 1979. – 600 с.
4. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 3-е изд. / Олифер В.Г., Олифер Н.А. – СПб: Питер, 2006. – 958 с.: ил. ISBN 978-5-469-00504-9. (Глава 7 – Методы обеспечения качества обслуживания).
5. 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.
6. Вишнеvский В.М., Семенова О.В. Системы поллинга: теория и применение в широкополосных беспроводных сетях. – М.: Техносфера, 2007. – 312 с. ISBN 978-5-94836-166-6.
7. Столингс В. Современные компьютерные сети. – СПб.: Питер, 2003. – 783 с.: ил. ISBN 978-5-94723-327-4. (Часть 3 – Моделирование и оценка производительности)
8. Авен О.И., Гурин Н.Н., Коган Я.А. Оценка качества и оптимизация вычислительных систем. - М.: Наука, 1982. - 464 с.
9. Феррари Д. Оценка производительности вычислительных систем. - М.: Мир, 1981.