

681.5
Д-33
А.А. ДЕНИСОВ · Д.Н. КОЛЕСНИКОВ

ТЕОРИЯ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

для студентов вузов



А. А. Денисов
Д. Н. Колесников

ТЕОРИЯ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

ДОПУЩЕНО МИНИСТЕРСТВОМ ВЫСШЕГО И СРЕДНЕГО
СПЕЦИАЛЬНОГО ОБРАЗОВАНИЯ СССР
В КАЧЕСТВЕ УЧЕБНОГО ПОСОБИЯ ДЛЯ СТУДЕНТОВ ВУЗОВ,
ОБУЧАЮЩИХСЯ ПО СПЕЦИАЛЬНОСТИ
«АВТОМАТИКА И ТЕЛЕМЕХАНИКА»



ЛЕНИНГРАД
ЭНЕРГОИЗДАТ
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ
1982

ББК 32.965
Д 33
УДК 681.51.01.(075.8)

681.5
Д-33

ПРЕДИСЛОВИЕ

70887

Рецензенты: кафедра «Автоматика и процессы управления»
ЛЭТИ им. В. И. Ульянова (Ленина) —
зав. кафедрой А. А. Вавилов; В. К. Дедков



Денисов А. А., Колесников Д. Н.

Д 33 Теория больших систем управления: Учебное пособие для вузов. — Л.: Энергоиздат, Ленингр. отд-ние, 1982. — 288 с., ил.

В пер.: 85 к.

В книге излагаются основы теории больших систем управления (БСУ), включающие в себя основные принципы системного подхода к изучению БСУ, вопросы формализации задач структурного анализа, аналитические и имитационные способы исследования поведения систем и анализ их характеристик с использованием моделей систем массового обслуживания. Излагаются также новые методы исследования БСУ, основанные на информационном подходе.

Книга предназначена в качестве учебного пособия для студентов, обучающихся по специальности «Автоматика и телемеханика»; она может быть полезна также специалистам в области теории систем обработки информации, автоматизированного управления.

Д 1502000000—053
051(01)—82 198—82

ББК 32,965
6Ф6.5

© Энергоиздат, 1982

681.5.01 + 519.86

Данная книга является учебным пособием по курсу «Большие системы управления» для студентов специальности «Автоматика и телемеханика» и может быть рекомендована также студентам других специальностей, изучающим вопросы построения любых сложных автоматизированных систем обработки информации и управления. В основу книги положен курс лекций, читаемых авторами на факультете автоматизации управления Ленинградского политехнического института им. М. И. Калинина. Трудности написания данного учебного пособия во многом определялись следующими причинами: 1) в настоящее время литература по теории больших систем управления (БСУ) представлена многочисленными статьями, монографиями, которые затрагивают лишь отдельные стороны проблематики БСУ, не давая целостного, связного о них представления; 2) сама теория БСУ находится в стадии интенсивного развития и далека от своего завершения; 3) круг задач, относящихся к этой теории, очень разнообразен и различается как содержанием, так и методами решения.

В связи с этим при написании данного учебного пособия авторы стремились отобрать те достижения современной теории БСУ, которые, с одной стороны, в наибольшей степени соответствуют утвержденной программе курса, а с другой — получили широкое практическое применение при решении различных системотехнических задач, возникающих при проектировании БСУ. Одновременно с этим в учебном пособии излагаются новые методы, основанные на информационном подходе к анализу БСУ с целью выработки у будущих специалистов единой универсальной концепции управления, вытекающей из основных положений материалистической диалектики. При изложении материала предполагается, что студенты знакомы с содер-

жанием курсов «Теория автоматического управления», «Цифровая вычислительная техника», «Телемеханика», «Исследование операций». Ориентация на системный подход к изучению БСУ определила следующую последовательность изложения:

1) в гл. 1 дается понятие БСУ и рассматриваются основные определения и принципы системного подхода к ее изучению;

2) гл. 2 посвящена определению основных понятий, вводимых при решении задач структурного анализа и формализации задач на различных уровнях задания исходной информации о БСУ;

3) в гл. 3 рассматривается подход к созданию общих моделей функционирования БСУ и исследуются особенности их поведения с учетом управления;

4) в гл. 4, 5 исследуются возможные способы описания и анализа поведения БСУ при наличии случайных факторов; эти способы получили широкое распространение при аналитическом подходе к определению различных функциональных характеристик как БСУ в целом, так и ее подсистем;

5) гл. 6 посвящена изложению общей методологии имитационного моделирования как способа исследования БСУ на ЭВМ в широком диапазоне условий;

6) в гл. 7 предлагается новый общий информационный подход к анализу поведения БСУ и исследуются его возможности;

7) гл. 8 содержит основные необходимые сведения о постановке и формализации задач структурного синтеза и методах их решения.

Все излагаемые основные понятия и методы иллюстрируются большим числом примеров.

Авторы, безусловно, не претендуют на полноту изложения всех вопросов теории БСУ в силу ограниченности объема данного учебного пособия, а делают лишь попытку представить основные, наиболее важные (с их точки зрения) разделы современной теории больших систем автоматизированного управления.

Все замечания и пожелания по книге просим направлять по адресу: 191041, Ленинград, Д-41, Марсово поле, 1, Ленинградское отделение Энергоиздата.

Авторы

СИСТЕМНЫЙ ПОДХОД К ИЗУЧЕНИЮ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

1-1. ПОНЯТИЕ БСУ

Одной из характерных тенденций развития общества в настоящее время является появление больших чрезвычайно сложных систем. Основными причинами этого являются: непрерывно увеличивающаяся сложность технических средств, применяемых в народном хозяйстве; необходимость в повышении качества управления как техническими, так и организационными объектами (предприятие, отрасль, государство и др.); расширяющаяся специализация и кооперирование предприятий — основные тенденции развития нашего народного хозяйства.

В отличие от традиционной практики проектирования несложных систем при разработке крупных автоматизированных, технологических, энергетических, гидротехнических, информационных и других сложных комплексов возникают проблемы, меньше связанные с рассмотрением свойств и законов функционирования элементов, а больше — с выбором наилучшей структуры, оптимальной организации взаимодействия элементов, определением оптимальных режимов функционирования, учетом влияния внешней среды и т. д.

По мере увеличения сложности системы этим комплексным общесистемным вопросам отводится все более значительное место. Понятие БСУ в значительной степени перекликается с понятием сложной системы. Однако в настоящее время для обоих понятий не существует строгих определений. Для таких определений необходимо найти четкие формальные признаки либо параметры, по которым можно было бы отличить большую систему от небольшой, сложную — от простой. Рассмотрим кратко некоторые подходы к определению данных понятий, существующие на сегодняшний день.

В качестве формальных признаков при определении сложной системы выделяют 1) число взаимосвязанных

элементов; 2) отсутствие формальной математической модели функционирования; 3) способ описания.

Отсюда и возникают различные определения. Советский ученый Г. Н. Поваров в зависимости от числа элементов, входящих в систему, выделяет четыре класса систем: малые системы ($10-10^3$ элементов), сложные (10^4-10^7 элементов), ультрасложные (10^7-10^{30} элементов), суперсистемы ($10^{30}-10^{200}$ элементов). Так как понятие элемента возникает относительно задачи и цели исследования системы, то и данное определение сложности является относительным, а не абсолютным.

Английский кибернетик С. Бир классифицирует все кибернетические системы на простые и сложные в зависимости от способа описания: детерминированного или теоретико-вероятностного, а советский ученый А. И. Берг определяет сложную систему как систему, которую можно описать не менее чем на двух различных математических языках (например, с помощью теории дифференциальных уравнений и алгебры Буля).

Очень часто сложными системами называют системы, которые нельзя корректно описать математически либо потому, что в системе имеется очень большое число элементов, неизвестным образом связанных друг с другом, либо неизвестна природа явлений, протекающих в системе. Все это свидетельствует об отсутствии единого определения сложности системы.

Аналогичная ситуация имеет место и в определении понятия БСУ. В ряде работ в качестве отличительного признака предлагается выбрать иерархичность структуры. Несмотря на некоторую убедительность, это определение обладает тем недостатком, что оно связывает понятие БСУ только с одной ее характеристикой — структурой и, что самое главное, только с одним видом структуры — иерархической. В других работах под БСУ понимают систему управления, в которой имеет место взаимодействие человека и машины (система «человек — машина»). Однако при таком определении в класс БСУ не попадают сложные, полностью автоматически управляющие комплексы. В некоторых случаях понятие БСУ связывают с неопределенностью ее поведения, вызванной большим числом случайных воздействий и участием в системе людей.

Во многих работах в качестве существенных признаков БСУ выделяют наличие больших потоков информации, сложной информационной структуры и достаточно сложных

алгоритмов переработки информации. Данное определение хотя и не является четким, однако верно отражает изменения, происшедшие в теории и практике создания систем управления.

С появлением и развитием автоматизированных систем управления (АСУ) в ряде работ, например в [9], понятие БСУ отождествляется с понятием АСУ. В этом случае под БСУ понимают совокупность материальных и человеческих ресурсов (сырье, материалы, полуфабрикаты, денежные средства, различные виды энергии, станки, оборудование, люди, занятые на выпуске продукции), средств преобразования, передачи и обработки информации, операторов, занятых на обслуживании этих средств, и руководителей, облеченных надлежащими правами и ответственностью для принятия решений, объединенных с помощью некоторой системы связей для достижения общей цели или группы целей.

Такой подход в определении значительно сужает класс БСУ (исключаются транспортные и телефонные сети больших городов и др.). Поэтому, как отмечено в [6], в силу отсутствия четкого определения отнесение системы к разряду больших является в значительной мере условным и связано в основном с тем, какую роль играют при изучении комплексные общесистемные вопросы. Это обстоятельство зависит как от свойств самих систем, так и от тех задач, ради решения которых проектируются системы. Рассмотрим ряд типичных примеров систем, которые могут быть отнесены к классу БСУ.

Автоматизированные системы управления. Всякая система управления с точки зрения ее функционирования решает три основные задачи: сбор и передача информации об управляемом объекте, переработка информации и выдача управляющих воздействий на объект управления.

Различают два основных типа автоматизированных систем управления (АСУ): системы организационно-экономического или административного управления и системы управления технологическими процессами (АСУ ТП). Главное отличие этих двух типов систем заключается прежде всего в характере объекта управления. Для первых объектом управления являются предприятие (АСУП), отрасль народного хозяйства (ОАСУ), т. е. человеческие коллективы, для вторых — это различные машины, процессы, приборы, устройства. Другое отличие — форма передачи информации. В АСУ ТП основной формой передачи информации являются различные сигналы (электрические, световые, механические и др.), в системах же организационно-экономического управления основная форма передачи информации — документ.

В настоящее время наметилась тенденция слияния АСУП и АСУ ТП в единые интегрированные системы управления; тем самым грани

между АСУ ТП и АСУП, до известной степени, стираются. Рассмотрим основные особенности и свойства АСУ как большой системы управления на примере АСУП (рис. 1-1).

АСУП — системы управления производственно-хозяйственной деятельностью предприятия, органически включающие в себя интегрированные системы обработки данных, главной целью которых является автоматизация информационных процессов на предприятии и усовершенствование формы и организации их выполнения. Объект управления — совокупность процессов, свойственных данному предприятию, по преобразованию ресурсов (материалы, полуфабрикаты, инструменты, оснастка, оборудование, энергетические, трудовые и финансовые ресурсы) в готовую продукцию. Сложность управления в АСУП обусловлена многими причинами: 1) большим числом разнородных элементов;

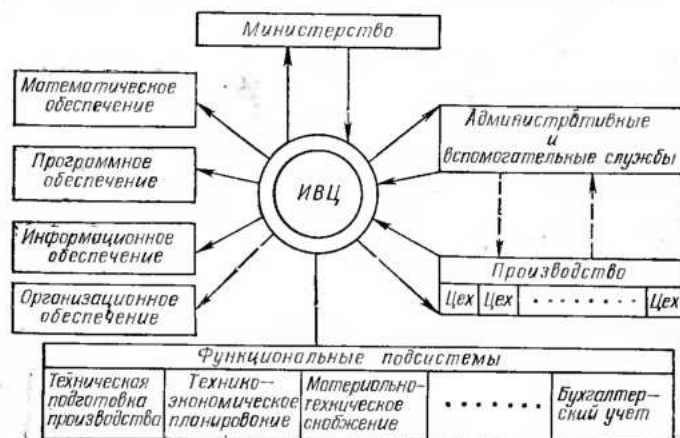


Рис. 1-1. Структурная схема АСУП

2) высокой степенью их взаимосвязи в процессе производства; 3) неопределенностью результатов выполнения многих процессов (брак, сбой, несвоевременные поставки, нерегулярность спроса); 4) тем, что объектами и субъектами управления на предприятии являются люди, а управление их поведением не столь очевидно и прямолинейно; 5) тем, что предприятие постоянно изменяется, развивается, т. е. является нестационарным.

Создание и внедрение АСУП привело к тому, что информационным процессам, их организации, проектированию, подготовке и выполнению уделяется такое же внимание, как и производственным. В структуре управления предприятием возникает специальное подразделение — информационно-вычислительный центр (ИВЦ), ответственный за упорядочение, регламентацию и непосредственное выполнение информационных процессов на предприятии. В структуре АСУП обычно выделяют функциональные и обеспечивающие подсистемы. Функциональные подсистемы непосредственно выполняют функции управления производственно-хозяйственной деятельностью предприятия. Такими функциями являются: конструкторско-технологическая подготовка производства, техническая подготовка производства, снаб-

жение, комплектация, складирование, планирование производства (основного, вспомогательного), сбыт, реализация продукции, финансовые операции и бухгалтерский учет, экономический анализ производственно-хозяйственной деятельности, учет кадров.

В зависимости от сложности управления той или иной функцией в АСУП выделяется та или иная подсистема, например подсистема оперативного планирования и диспетчеризации основного производства. Обеспечивающие подсистемы выполняют собственно все информационные процессы в АСУП и ответственны за их подготовку и организацию. Чаще всего выделяют: подсистемы математического, программного, технического, информационного, организационного обеспечения.

Математическое обеспечение — модели, методы, алгоритмы и их обоснование для решения задач управления в соответствующих функциональных подсистемах и выполнения соответствующих информационных процессов в АСУП.

Программное обеспечение включает в себя комплекс программ и инструкций к ним для решения задач. Его обычно разделяют на специальное и общее. Специальное программное обеспечение нацелено на получение документов и сведений производственно-хозяйственного значения. Общее программное обеспечение включает в себя программы, предназначенные для различных преобразований над массивами данных системы.

Техническое обеспечение — технические средства для механизации и автоматизации выполнения основных информационных процессов (сбор, передача, обработка информации), а также инструкции по их эксплуатации и обеспечению надежного функционирования.

Информационное обеспечение регламентирует потоки и подготовку информации, организацию и выполнение информационных процессов в ИВЦ.

Организационное обеспечение регламентирует действие каждого работника управления, каждого рабочего по отношению к системе информации и всей схеме принятия решений в АСУП.

Наиболее полно все перечисленные элементы АСУП представлены, например, в системе «Львов» для телевизионного завода. В этой системе все информационные процессы, включая фиксацию и подготовку первичной информации, отражающей ход и состояние производственно-хозяйственной деятельности, обработку этой информации и подготовку различной отчетности, сосредоточиваются в координационно-управляющем центре (КУЦ) предприятия.

Координационно-управляющий центр, построенный на базе двух ЭВМ «Минск-22», работает в замкнутом контуре управления предприятием в реальном масштабе времени. Схемно-программные средства позволили организовать режим разделения времени для совмещения непрерывного обмена оперативной информацией с процессами решения основных задач управления и обработки данных.

Другие технические средства, разработанные в системе «Львов», обеспечивают дистанционный ввод в КУЦ оперативно-производственной информации от различных источников непосредственно в момент ее возникновения, а также вывод необходимых сообщений в разные производственные подразделения и службы аппарата управления предприятием. Взаимосвязанно с вычислительным комплексом работают диспетчерские пульта завода и цехов, табло и счетчики для визуального наблюдения за параметрами, определяющими общую динамику про-

изводства; они специально разработаны для системы «Львов». Внедрение системы «Львов»: 1) привело к сокращению запаздывания между отклонением от нормального запланированного хода производства, с одной стороны, и принятием решений по формированию управления — с другой; 2) позволило решить новые задачи по управлению предприятием, прежде не решавшиеся из-за их трудоемкости, такие, как оперативно-календарное планирование заготовительных цехов, определение размера партий деталей и периодичности их запуска, выдача цехам оперативных планов производства с учетом наличия деталей в незавершенном производстве; 3) позволило по-новому организовать работу управляющего аппарата; труд его более четко проектируется, планируется, контролируется; человек становится неотъемлемым звеном в человеко-машинной системе управления; 4) привело к значительному росту эффективности управления предприятием, обеспечило дополнительное увеличение выпуска продукции на 7%, снижение уровня запасов на 20%, ускорение оборачиваемости оборотных средств на 10%, сокращение инженерно-технического и административно-управленческого персонала.

Экономическая эффективность системы составила около полумиллиона рублей в год, срок ее окупаемости — 1 год.

При проектировании подобных АСУП большая работа должна быть выполнена при предсистемном обследовании объекта управления. При этом должны быть решены следующие задачи: 1) анализ организационной структуры предприятия — изучение основных технико-экономических характеристик предприятия и его организационной структуры для выработки рекомендаций по совершенствованию последней; 2) анализ функциональной структуры предприятия — изучение функций управления отдельных должностных лиц в структурных подразделениях и функциональных подсистемах для выработки рекомендаций по упорядочению последовательности их выполнения, содержания функций управления, их границ и взаимосвязей; 3) анализ потоков и состава информации между подразделениями и внутри их для выработки рекомендаций по устранению избыточности информации в звеньях управляющей части системы, рационализации информационных потоков и структуры документированной информации; 4) анализ материальных потоков предприятия для выработки рекомендаций по их совершенствованию и упорядочению, а также выявлению особенностей потоков, влияющих на характер АСУП; 5) анализ методов планирования и учета — изучение применяемых на предприятии методов планирования и учета производства для выбора способов реализации их в АСУП; 6) выработка общих рекомендаций по улучшению управления (осуществляется по итогам анализа результатов обследования), в том числе без дополнительного применения средств автоматизации и электронно-вычислительной техники; 7) выбор объектов автоматизации — состава автоматизируемых функций и задач управления по подразделениям, а также технико-экономических показателей, отражающих влияние автоматизации; 8) предварительный выбор технических средств АСУП, необходимых для автоматизации выбранных процессов управления; 9) разработка укрупненного перечня наиболее трудоемких мероприятий по подготовке объекта к внедрению АСУП; 10) определение намечаемых затрат и основных технико-экономических показателей, которые будут достигнуты в результате создания АСУП.

Решение указанных задач позволяет разработать техническое задание на разработку АСУП и обосновать эффективность такой разработки. Обоснованное и подробное техническое задание является осно-

вой для успешного развертывания широкого фронта работ по созданию математического, программного, информационного, технического, организационного обеспечений, причем не в отрыве друг от друга, а в полном единстве в составе разрабатываемой АСУП.

Из характера задач при проектировании АСУП видно, что в ее создании принимает участие большой коллектив различных специалистов: математиков, программистов, специалистов по информационной и вычислительной технике и др. Поэтому успешное завершение проекта и внедрение АСУП зависит от правильного руководства на всех стадиях проекта.

Автоматизированные системы управления обслуживанием. Широкое распространение в настоящее время получили автоматизированные системы управления обслуживанием на авиационном, железнодорожном транспорте и др. В литературе [46] подобные системы получили название телеавтоматических систем массового обслуживания (ТАСМО). Одним из характерных примеров такой системы является система «Сирена» резервирования мест на самолеты.

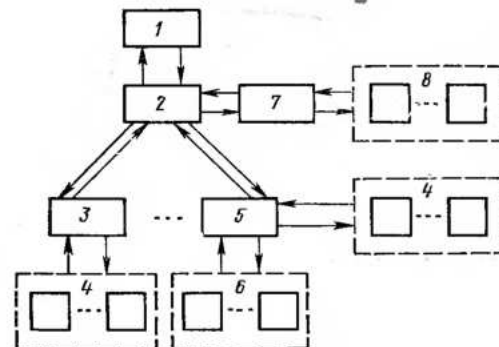


Рис. 1-2. Структурная схема телеавтоматической системы массового обслуживания типа «Сирена»

Цель введения такой системы: более оперативное обслуживание пассажиров, повышение загрузки самолетов, совершенствование контроля операций, выполняемых в течение дня, сбор статистических сведений для решения задач перспективного планирования. Сложность управления обслуживанием в этой системе определяется следующими причинами: объект обслуживания — большая территориально-разветвленная сеть возможных источников запросов, причем источники не являются однородными; число источников колеблется в пределах от нескольких сотен до нескольких тысяч; случайный характер требований на данные и программы; нестационарность потоков запросов в течение суток; управление в реальном масштабе времени, причем с малым временем реакции на запрос (2,5—3 с); круглосуточная работа; охват контуром человека.

На структурной схеме системы «Сирена» (рис. 1-2) приняты следующие обозначения: 1 — центр обработки данных; 2 — устройство ввода-вывода, обеспечивающее связь как с абонентом в пределах данного города, так и с иногородними абонентами через АТС либо через

телеграфные каналы связи; 3 — станция абонентного телеграфирования; 4 — набор телетайпов; 5 — блок коммутации сообщений, повышающий эффективность передачи и уменьшающий задержки информации; 6 — пульты-мониторы, которые устанавливаются у кассиров; 7 — устройство согласования, обеспечивающее согласование с иногородними каналами связи и иногородними абонентами; 8 — периферийные пункты, устанавливаемые в других городах. Пульты-мониторы обеспечивают оперативный ввод информации, регистрацию и одновременно визуальный вывод информации непосредственно на пульт.

У крупленной структуры подобных систем обычно включает в себя: 1) систему терминалов для общения кассиров и авиационных агентов с системой; 2) средства связи; 3) центральный вычислительный комплекс для обработки данных. При их проектировании необходимо обеспечить малое время реакции на запрос, высокую достоверность обработки и живучесть системы. Поэтому рабочему проектированию системы должны предшествовать многочисленные предварительные исследования: 1) по анализу входных потоков информации для определения объемов, длин сообщений, распределению размеров и типов сообщений, а также числа обращений к различным программам; 2) по выбору соответствующих методов уплотнения в каналах связи, определению необходимой степени концентрации потоков и режимов опроса терминалов; 3) по обоснованию структуры и режимов работы центра обработки данных; 4) по разработке эффективной операционной системы для обеспечения малого времени реакции и сохранения стоимости системы в заданных пределах; 5) по созданию банка данных с малым временем доступа; 6) по обеспечению адаптивности системы в режимах и в структуре при отклонении информационной нагрузки от нормальной, при возникновении отказов и сбоев.

Вычислительный комплекс системы «Сирена» (рис. 1-3) построен на базе агрегатной системы средств вычислительной техники АСВТ-Д. Процессор системы выполняет арифметическую и логическую обработку информации и является универсальным устройством. Он может иметь переменный состав. При минимальной комплектации он содержит блоки обработки команд 1, микропрограммного управления 3 и обработки информации с фиксированной запятой 2, а также инженерный пульт 7. В расширенном составе, кроме упомянутых устройств, в состав процессора входят блоки внутренней памяти 6, обработки информации с плавающей запятой 4, обработки символьной и десятичной информации 5.

Быстродействие процессора в максимальном составе — 100 000 операций типа сложения с фиксированной запятой. Оперативная и постоянная память процессора состоит из 11 модулей по 1536 36-разрядных слов. При необходимости иметь более 11 модулей используется устройство наращивания памяти (УНП). Входящие в состав комплекса устройства защиты памяти (УЗП) защищают хранящуюся здесь информацию от ошибочной записи посторонних данных.

Каждый мультиплексный канал (УМС) обеспечивает подключение до 256 низкоскоростных устройств ввода-вывода (алфавитно-цифровых печатающих устройств, устройств ввода с перфокарт). Селекторные каналы используются для подключения устройств внешней памяти. Каналы связи к процессорам подключаются через устройства сопряжения машин (УСМ) и многоканальные адаптеры (МАД). Один МАД обеспечивает подключение до 32 каналов связи как закрепленных, так и коммутируемых. К одному УСМ можно подключать до четырех МАД.

Вычислительный комплекс обладает избыточной структурой, обеспечивающей его высокую живучесть: 1) в системе предусмотрено резервирование процессоров и оперативных и постоянных 3У; наличие в вычислительном комплексе внутрисистемной связи (УС) обеспечивает доступ процессора ВК1 к оперативным и постоянным 3У ВК2; 2) при необходимости устройства ввода-вывода ВК1 через межканальный адаптер МКА могут быть подключены к ВК2. В силу этого возможны различные сочетания работающих устройств: работа процессора ВК1 с оперативными 3У комплексов ВК1, ВК2; работа двух процессоров ВК1, ВК2 с 3У либо ВК1, либо ВК2; работа двух процессоров и одного комплекта устройств ввода-вывода.

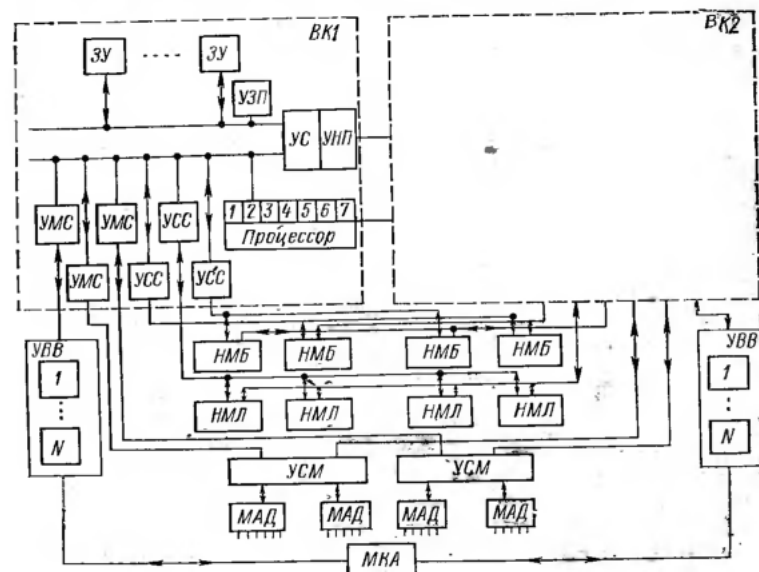


Рис. 1-3. Структурная схема вычислительного комплекса АСУ «Сирена»

ВК1, ВК2 — отдельные вычислительные комплексы; 3У — запоминающее устройство; УЗП — устройство защиты памяти; УНП — устройство наращивания памяти; УМС — устройство мультиплексной связи; УСС — устройство селекторной связи; УВВ — устройство ввода-вывода; УС — устройство внутрисистемной связи; УМБ — накопитель на магнитном ленте; УМЛ — накопитель на магнитном барабане; УСМ — устройство сопряжения машин; МКА — межканальный адаптер; МАД — многоканальный адаптер

Вычислительный комплекс может иметь: нормальный режим асинхронного дуплекса (каждый процессор обрабатывает одни и те же требования от абонентов независимо и с произвольным сдвигом во времени), который обеспечивает повышение достоверности; симплексный режим с повторным счетом в случае отказа одного из процессоров; мультисистемный — режим параллельного выполнения процесса обработки с общей диспетчеризацией для повышения производительности системы в часы пик нагрузки.

Наличие избыточной структуры и многорежимного характера работ вычислительного комплекса требует сложной организации взаимодействия элементов комплекса в процессе вычислений на основе операционной системы, являющейся центральным ядром математического обеспечения. Изложенное выше наглядно показывает, что не только сама система автоматизированного управления обслуживанием, но и ее подсистемы, такие, как вычислительный комплекс, обладают функциональной и структурной сложностью.

Автоматизированная система управления обслуживанием «Сирена» осуществляет распределение мест на самолеты, контролируя места на 1100 рейсов в пределах тридцати дней предварительной продажи, и рассчитана на обработку до 250 000 запросов в течение суток. Среднее время реакции системы, т. е. среднее время ответа на запрос кассира о наличии мест в данном самолете данного рейса, составляет 3 с. Время автоматического печатания билета составляет 2,5 с.

Автоматизированные системы контроля. К разряду БСУ могут быть отнесены и современные автоматизированные системы контроля (АСК) сложных объектов, получившие широкое развитие в настоящее время [31]. Они предназначены для проверки оборудования радиолокационных комплексов, автоматизированных систем управления самолетами, ракет, кораблей флота и других сложных технических объектов, применяемых в народном хозяйстве.

Цель создания АСК включает: сокращение общих трудозатрат на техническое обслуживание, уменьшение времени подготовки техники к выполнению задачи, повышение достоверности результатов контроля, повышение коэффициента использования технических ресурсов объектов контроля, предупреждение крупных повреждений и разрушений технических объектов. Современная АСК есть автоматизированная система измерения и обработки информации. Она предназначена для контроля сложных объектов и представляет собой совокупность программно-управляемых технических средств (измерительных, вычислительных, вспомогательных), имеющих блочно-модульную структуру, определенную организацию и связи, которые обеспечивают получение, преобразование, накопление, обработку и выдачу измерительной, командной и другой информации в соответствующей форме, в том числе для воздействия и на объект исследования. Обобщенная структура АСК приведена на рис. 1-4.

Структурная и функциональная сложность построения АСК в основном определяется: 1) сложностью современных объектов контроля (число контролируемых параметров колеблется в пределах от 100 до 1000); 2) необходимостью организации контроля в реальном масштабе времени; 3) наличием большого числа случайных факторов, присущих как объектам контроля, так и самой АСК (в каналах измерения и обработки информации возникают помехи и сбой); 4) высокой степенью автоматизации процессов сбора, передачи, обработки и представления результатов оператору для принятия окончательного решения; 5) жесткими требованиями, предъявляемыми к основным качественным характеристикам АСК: достоверности и быстродействию.

Достоверность — свойство системы обеспечивать соответствие результата контроля истинному состоянию проверяемого оборудования. Достоверность контроля зависит от целого ряда факторов, влияющих на методические и инструментальные составляющие достоверности.

Методическая составляющая достоверности контроля обуславливается следующими основными факторами:

1) законами распределения значений выходных параметров контролируемого объекта и погрешностей измерений, набором допусков работоспособности; 2) набором контрольных допусков, взятых за основу принятия решений в АСК: «годен», «негоден»; 3) набором контролируемых параметров, определяющих глубину (полноту) контроля; 4) периодом проверки объекта и временем контроля этого объекта.

Инструментальная составляющая достоверности контроля обуславливается следующими факторами, к которым относятся: 1) надежность АСК (влияние на достоверность отказов, сбоев); 2) методы организации самоконтроля АСК; 3) способы накопления, регистрации и отображения результатов контроля.

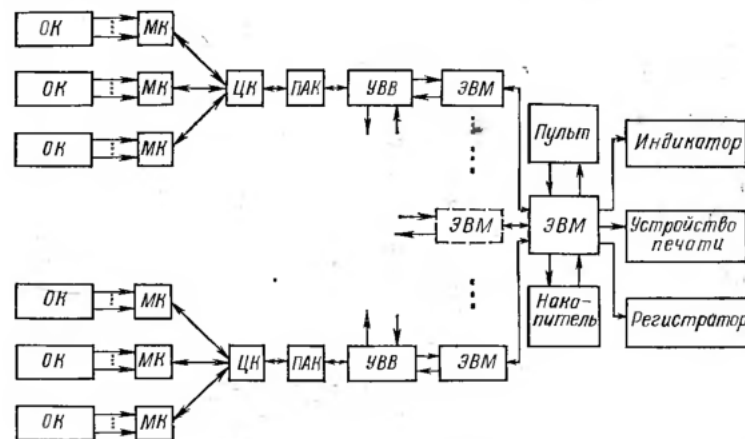


Рис. 1-4. Структура рассредоточенной системы контроля бортового оборудования крупного самолета

ОК — объект контроля; МК — местный коммутатор; ЦК — центральный коммутатор; ПАК — преобразователь «аналог — код»; УВВ — устройство ввода-вывода

Быстродействие контроля зависит также от целого ряда факторов, основными из которых являются [31]: число контролируемых параметров, производительность системы обработки информации в АСК, надежность АСК, методы организации самоконтроля и др. Обычно в качестве оценки быстродействия выбирается среднее время выполнения контроля объекта. Достоверность чаще всего оценивается средней вероятностью принятия ошибочного решения. Среднее время контроля сложных объектов колеблется от 10 мин до 1,5 ч, вероятность ошибочного решения — от 10^{-2} до 10^{-6} .

Для оценки степени соответствия основных показателей качества проектируемой АСК требуемым на этапе проектирования необходимо решить целый ряд комплексных системных задач: 1) выбрать и унифицировать общую систему измерительных средств с учетом всего многообразия выходных параметров объекта контроля; 2) разработать общую методику оценки методической достоверности контроля; 3) провести сбор статистических данных о законах распределения значений вы-

ходных параметров и погрешностей измерений; 4) провести количественный анализ методической достоверности с учетом принятого алгоритма контроля и выбранной системы средств измерений; 5) выбрать и обосновать количественно структуру подсистемы сбора информации об измеряемых параметрах, алгоритм опроса и другие параметры; 6) выбрать и обосновать количественно структуру и режим работы вычислительных средств в АСК (однопроцессорная, многопроцессорная, последовательный режим, мультипрограммный режим), исходя из необходимых требований к быстродействию контроля; 7) выбрать подсистему самоконтроля в АСК; 8) провести общий количественный анализ показателей быстродействия и достоверности с учетом всего множества указанных выше факторов; 9) определить целесообразную степень автоматизации и др.

Процесс проектирования современных сложных АСК занимает 4—5 лет и требует очень детального предварительного исследования, направленного на проверку возможности создания системы с необходимыми качественными и количественными характеристиками.

Типичными представителями реальных объектов, которые могут рассматриваться как БСУ, являются телефонные сети крупных городов, энергетические комплексы, гидротехнические сооружения, экономические, биологические системы и т. д.

Приведенных примеров достаточно для того, чтобы подчеркнуть некоторые общие свойства и отличительные особенности БСУ: 1) наличие большого числа взаимосвязанных и взаимодействующих элементов, причем изменение в характере функционирования какого-либо из элементов отражается на характере функционирования других и всей системы в целом; 2) система, а также входящие в нее разнообразные элементы в подавляющем большинстве являются многофункциональными; 3) взаимодействие элементов в системе может происходить по каналам обмена информацией, энергией, материалами и др.; 4) наличие у всей системы общей цели, общего назначения, определяющего единство сложности и организованности, несмотря на все разнообразие входящих в нее элементов; 5) переменность структуры (связей и состава системы), обеспечивающей многорежимный характер функционирования, возможность адаптации как в структуре, так и в алгоритме функционирования; 6) взаимодействие элементов в системе и с внешней средой в большинстве случаев носит стохастический характер; 7) система является эрготической, ибо часть функций всегда выполняется автоматически, а другая часть — человеком; при этом следует отметить высокую степень автоматизации, в частности широкое применение средств автоматики и вычислительной техники для гибкого управления и механизации умственного и ручного труда человека, работающего в системе; 8) управление в подавляю-

щем большинстве систем носит иерархический характер, предусматривающий сочетание централизованного управления или контроля с автономностью ее частей.

Следует, однако, отметить, что для включения рассматриваемой системы в класс БСУ совсем не является обязательным наличие полного перечня признаков, перечисленных выше. Рассмотренные выше особенности БСУ, несмотря на отсутствие единого их определения, привели к необходимости разработки новых подходов к изучению подобных систем, объединенных общим названием — системный подход.

1-2. СИСТЕМНЫЙ ПОДХОД. ОСНОВНЫЕ ЗАДАЧИ СИСТЕМОТЕХНИКИ

Для рассмотрения основных задач системного подхода введем следующие определения и понятия, принятые в общей теории систем:

А. Система есть прежде всего совокупность элементов, которые, в свою очередь, при определенных условиях могут рассматриваться как системы, а сама исследуемая система — как один из элементов более широкой системы.

Б. Для любых систем характерно наличие и н т е г р а т и в н ы х к а ч е с т в (свойств). Интегративными называются качества, присущие системе в целом, но не свойственные ни одному из ее элементов в отдельности. Отсюда следует важный вывод: система не сводится к простой совокупности элементов, и, расчленив систему на отдельные части, изучая каждую из них в отдельности, нельзя познать все свойства системы в целом.

В. Для любых систем характерно наличие с у щ е с т в е н н ы х с в я з е й между элементами, превосходящих по мощности (силе) связи этих элементов с элементами, не входящими в данную систему. Указанное свойство позволяет выделить систему в виде целостного объекта из окружающей среды.

Изучение любой системы предполагает создание модели системы, позволяющей предсказывать ее поведение в определенном диапазоне условий. Модель — описание системы, отображающее определенную группу ее свойств; углубление описания — детализация модели. Описание системы можно рассматривать с трех точек зрения: а) функциональной; б) морфологической; в) информационной.

Функциональное описание необходимо для того, чтобы осознать важность системы, определить ее место, оценить отношение к другим системам и к внешней среде. Для построения общего функционального описания системы вводятся следующие понятия.

Состояние — множество существенных свойств, которыми система обладает в данный момент времени. **Внешняя среда** — множество элементов, которые не входят в систему, но изменение их состояния вызывает изменение состояния системы. **Модель функционирования** (поведения) системы — модель, предсказывающая изменение состояния системы во времени.

Морфологическое описание должно дать представление о строении системы. Оно не может быть исчерпывающим; глубина описания, уровень детализации, т. е. выбор элементов, внутрь которых описание не проникает, определяется назначением системы и целью исследования. Изучение морфологии начинается с элементарного состава. Под элементом в данном случае понимается подсистема, внутрь которой описание не проникает. Морфологические свойства системы существенно зависят от характера связей. Обычно выделяют: информационные, энергетические, вещественные связи. Информационные связи предназначены для переноса информации, энергетические — для переноса энергии между элементами, вещественные — для передачи материала, изменения свойств материала. В результате морфологического описания возникает понятие структуры. **Структура** — совокупность элементов и связей между ними.

Информационное описание системы должно давать представление об организации системы. Оно определяет зависимость морфологических и функциональных свойств системы от качества и количества внутренней (о себе самой и среде) и внешней (поступающей из среды) информации.

С общепhilософских позиций системный подход — направление методологии научного познания и социальной практики, в основе которого лежит исследование объектов как систем. Методологическая специфика системного подхода определяется тем, что ориентирует исследование на раскрытие целостности объекта и обеспечивающих ее механизмов, на выявление многообразных типов связей сложного объекта и сведение их в единую теоретическую картину. Впервые необходимость системного подхода к про-

ектированию БСУ была наглядно проиллюстрирована и обоснована американскими учеными Г. Х. Гудом и Р. Э. Макколом в их фундаментальной работе [14]. Они предлагают следующий пример: «Некто предлагает вам 100 000 000 долларов и говорит: «Сделайте мне систему для управления воздушным движением, или для запуска межконтинентальных баллистических ракет, или для соединения 7000 телефонных аппаратов, или для выполнения каких-либо других функций». Что вы при этом будете делать?»

Решение подобных задач как за один прием, так и при простом разбиении всей большой задачи на ряд мелких последовательных этапов редко приводит к успеху. Поэтому для создания таких систем необходим определенный идеологический и организационный план проектирования системы, пронизывающий весь процесс проектирования начиная от фазы исследовательской проработки до фазы изготовления, испытаний и применения опытного образца. Необходимость системного подхода при создании БСУ объясняется тем, что темпы развития науки и производства с каждым днем увеличиваются, сложность систем возрастает, а это увеличивает длительность разработки, в результате чего большинство систем к моменту ввода их в эксплуатацию могут оказаться морально устаревшими, а иногда и просто ненужными. Проектирование подобных систем требует больших капиталовложений, в результате чего требуются гарантии того, что будет создана система с нужными свойствами. Отсюда выявляется основное целевое назначение системного подхода к проектированию — сокращение периода проектирования системы между моментом возникновения потребности в создании подобной системы и моментом ввода ее в эксплуатацию при сохранении соответствия качества выполняемых функций требуемому для достижения поставленных целей.

Необходимость конструирования современных БСУ привела к активному использованию системного подхода в технике, в результате чего появилась новая научно-техническая дисциплина «системотехника», охватывающая вопросы проектирования сложных систем.

Согласно [14], термин «системотехника» используется для описания обобщенного подхода к проектированию систем, конструируемых для решения различных задач, при условии, что эти задачи будут решаться наиболее эффективным из всех возможных способом, причем каждый элемент си-

стемы будет рассматриваться не сам по себе, а во взаимодействии с другими элементами.

К числу задач, решаемых на основе системотехнического подхода, обычно относят: определение общей структуры системы, организацию взаимодействия между подсистемами и элементами, учет влияния внешней среды, выбор оптимальной структуры, оптимальных алгоритмов функционирования.

В монографии [6] предлагается все проектирование БСУ разделить на две стадии: макропроектирование (внешнее проектирование), в процессе которого решаются функционально-структурные вопросы системы в целом; микропроектирование (внутреннее проектирование), связанное с разработкой элементов системы как физических единиц оборудования и с получением технических решений по основным элементам (их конструкции и параметры, режимы эксплуатации).

В соответствии с предложенным делением системотехника определяется как совокупность точек зрения, методов и подходов, связанных с макропроектированием сложных систем.

Макропроектирование начинается с формулировки проблемы, которая включает в себя по крайней мере три основных раздела: 1) определение целей создания системы и круга решаемых ею задач; 2) описание действующих на систему факторов, подлежащих обязательному учету при разработке; 3) выбор показателя или системы показателей эффективности системы. Рассмотрим кратко характеристику каждого из разделов.

1. В начале проектирования необходимо правильно и четко определить главную цель разработки, которая преследуется при создании новой системы. Отказ от уяснения цели разработки не позволяет выполнить проектирование наилучшим образом: заставляет разработчиков при выборе возможных вариантов ориентироваться лишь на собственные цели, связанные, как правило, со стремлением минимизировать расход временных, материальных и людских ресурсов. В конечном итоге усилия большого коллектива людей и большие материальные затраты могут оказаться бесполезными и напрасными. Один из наиболее известных примеров неправильной постановки целей в истории развития системотехники является организация противовоздушной обороны английских торговых судов во время второй мировой войны. Исследуя эффективность такой меры, как

установка на торговые суда зенитных орудий, английские специалисты пришли сначала к выводу о необходимости отказа от нее. Причиной такого заключения было то, что в качестве цели первоначально была принята задача уничтожения зенитным огнем вражеских самолетов. Естественно, что зенитные орудия, стрелявшие с качающихся палуб и обслуживаемые недостаточно квалифицированными расчетами, почти не сбивали самолетов. Отсюда логическим образом следовало заключение о необходимости передачи зенитных средств с судов на наземные батареи, где эффективность их использования была бы значительно выше.

Специалисты, однако, вовремя поняли, что действительной целью установки зенитных орудий на торговые суда является не уничтожение вражеских самолетов, а защита самих судов. В этом смысле орудия оказались весьма эффективны: опасаясь зенитного огня, немецкие летчики бомбили вооруженные суда с больших высот и с гораздо меньшей точностью. Экономия от сокращения потерь судов намного перекрывала затраты на установку и обслуживание орудий.

В результате уяснения цели можно выявить единственную цель (как, например, увеличение пропускной способности информационно-справочной системы, либо минимизацию стоимости, либо минимизацию времени контроля сложного объекта) или несколько разных целей, которые преследуются одновременно. Так, например, в системе автоматизированного управления производством можно добиваться повышения производительности, снижения себестоимости и повышения качества продукции, а также уменьшения стоимости системы управления; в информационной системе «Сирена» резервирования билетов на самолеты — увеличения пропускной способности, уменьшения стоимости системы.

В общем случае следует отметить, что цели и задачи систем определяются исходя из потребностей практики, тенденций и особенностей развития современной техники, а также из народнохозяйственной, экономической целесообразности. В настоящее время нет возможности указать какие-либо формальные правила для решения этого вопроса, за исключением, может быть, понимания роли проектируемой системы, опыта применения существующих систем аналогичного типа и инженерной интуиции.

2. Необходимо четко оговорить внешние условия, в которых будет функционировать проектируемая система:

а) нужно выделить те связи с внешней средой, которые являются наиболее существенными и которые необходимо учитывать при проектировании; б) задать приближенное описание этих связей, для чего могут быть использованы как опыт эксплуатации аналогичных систем, так и статистические данные, полученные в результате специально поставленных экспериментальных исследований.

3. Под эффективностью понимается обобщенное свойство системы, характеризующее степень приспособленности системы к выполнению поставленных перед нею задач. Для количественной оценки эффективности выбирают один или несколько показателей (критериев). Характер выбранного показателя или системы показателей определяет основные направления в поиске свойств системы, обеспечивающих оптимальный характер выполнения поставленных перед нею задач. Проследим различные варианты выбора показателя эффективности и последствия этого выбора для различных систем.

Производственный процесс как объект автоматизации в АСУП. Выберем в качестве показателя эффективности управления производством производительность, измеряемую средним числом изделий, выпускаемых в течение заданного интервала времени. Исходя из такого показателя эффективности, при проведении мероприятий, связанных с усовершенствованием и управлением производством, будет большее внимание уделяться факторам, способствующим увеличению производительности. При этом неизбежно будут упущены из виду другие стороны организации производственного процесса, не связанные непосредственно с производительностью, такие, как качество изделий, экономия сырья, энергии и фонда заработной платы, предотвращение износа оборудования и др.

Предположим теперь, что в качестве показателя эффективности управления производством выбрана средняя себестоимость продукции. В этом случае такие факторы, как экономия сырья, износ оборудования, расход энергии и фонда заработной платы, будут иметь большой вес; вместе с тем отойдут на второй план факторы, связанные с производительностью оборудования и качеством продукции.

Автоматизированная система контроля сложного оборудования. В качестве основных показателей эффективности АСК могут быть выбраны показатели, характеризующие достоверность контроля. Напомним, что под достоверностью систем контроля понимается свойство системы, отражающее степень соответствия результатов контроля истинному состоянию объекта контроля. В этом случае при проектировании АСК основное внимание будет уделяться вопросам обеспечения максимальной достоверности, но не будет учитываться другое важное свойство системы — быстродействие. В результате могут быть выбраны средства повышения достоверности, далеко не лучшие по отношению к временным характеристикам процесса контроля.

Рассмотренные примеры убеждают в том, что выбор показателей эффективности необходимо проводить с боль-

шой осторожностью, после глубокого анализа целей и задач, поставленных перед системой. В общем случае выбранная система показателей эффективности зависит: от структуры системы и характера связей между элементами; от вида управляющих алгоритмов и закономерностей функционирования; от параметров внешней среды.

После формулировки проблемы обычно приступают к определению одного или нескольких вариантов построения системы, которые заслуживают дальнейшего рассмотрения. Исследование вариантов всегда начинается с анализа их эффективности.

Для получения методики расчета показателей эффективности необходимо построить адекватное математическое описание процесса функционирования системы или, другими словами, математическую модель, позволяющую выявить зависимость показателей эффективности от параметров системы и внешней среды, структуры и алгоритмов взаимодействия элементов в системе. Математическая модель является основой для решения главных системотехнических задач: анализа — определения численных значений показателей эффективности при заданных параметрах системы и характеристиках внешней среды, фиксированной структуре и алгоритме взаимодействия элементов; синтеза — выбора оптимальной структуры, алгоритмов взаимодействия, параметров системы, оптимального управления системой и др.

Зависимости между параметрами в БСУ являются разнообразными и сложными, в результате чего построение единой математической модели оказывается затруднительным. Поэтому для моделирования таких БСУ используют принцип многоуровневого (иерархического) описания [34]. Многоуровневое описание предполагает введение различных формальных языков, каждый из которых отражает функционирование системы в соответствии с понятиями и отношениями, принятыми на том или ином уровне иерархии. Для того чтобы получить такое математическое описание, необходимо выбрать параметры, позволяющие описывать функционирование системы: во-первых, как элемент более широкой системы; во-вторых, как целостное явление; в-третьих, как некоторую сложную систему, структуру которой необходимо представить с достаточной для данного уровня детализацией.

Применение многоуровневой модели вытекает из необходимости простого ее построения, в то же время необхо-

димому учесть многочисленные характеристики БСУ. Рассмотрим интерпретацию данного понятия применительно к большим автоматизированным системам управления.

Уровень 1 — информационное описание, соответствующее взгляду на систему в целом и на ее взаимодействие с внешней средой. При этом в первую очередь разработчиков должны интересовать все информационные связи системы с внешней средой и роль системы как некоторого преобразователя информации в ее информационном взаимодействии с внешней средой.

Уровень 2 — функциональное описание. Уровень функционального описания выявляет способ реализации закона управления и определяет множество функциональных элементов АСУ и отношений между ними. В результате вырисовывается функциональная структура системы; при этом каждая функциональная подсистема выполняет определенную часть общего алгоритма управления.

Уровень 3 — системотехническое описание. Уровень этого описания выявляет структуру комплекса технических средств АСУ, под которой понимается состав и связи функциональных групп оборудования (технических подсистем), номенклатура, число и размещение технических средств каждой функциональной группы. Технические подсистемы предназначены для реализации отдельных самостоятельных функций в составе общего процесса преобразования информации.

В соответствии с тремя уровнями описания возникают три группы системных задач, решаемых на этапе проектирования АСУ: определение взаимоотношений управляющей системы с внешней средой и объектом управления и формирование закона управления; алгоритмизация закона управления и разработка функциональной структуры; выбор технических средств для реализации информационных процессов и разработка структуры комплекса технических средств.

Как отмечено в [34], многоуровневое описание характеризуется целым рядом общих свойств.

1. Выбор уровня описания, в терминах которого описывается данная система, зависит главным образом от цели исследования. Для многих систем выбор уровня описания является естественным и определяется назначением системы. Выделение нескольких уровней для изучаемой системы позволяет вести параллельное построение моделей на каждом уровне различными специалистами.

2. Аспекты описания функционирования системы на различных уровнях в общем случае не связаны между собой, поэтому принципы и законы, используемые для описания системы на любом уровне, не могут быть формально выделены из принципов, используемых на других уровнях.

3. Существует асимметричная зависимость между условиями функционирования системы на различных уровнях, т. е. требования, предъявляемые к работе системы на верхнем уровне, выступают как условия или ограничения для нижестоящих уровней.

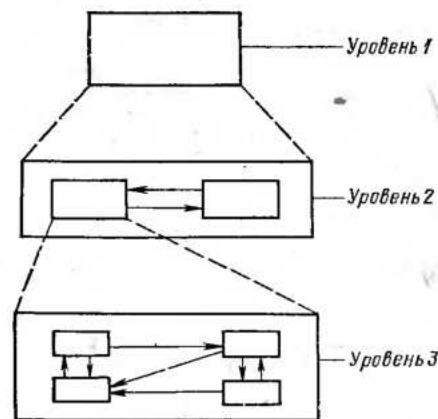


Рис. 1-5. Многоуровневое описание БСУ

4. На каждом уровне имеется свой собственный набор понятий и принципов, составляющих язык описания системы. Таким образом, в соответствии с иерархией уровней возникает иерархия языков описания БСУ (рис. 1-5).

5. Понимание системы возрастает при последовательном переходе от одного уровня к другому: чем ниже производится спуск по иерархической лестнице, тем более детально становится раскрытие системы, чем выше мы поднимаемся, тем яснее становится смысл и назначение всей системы. Решение системотехнических задач для БСУ возможно только на основе многоуровневого принципа построения моделей.

Совокупность математических методов для решения поставленных системотехнических задач анализа и синтеза на основе построенных моделей составляет содержание другой важной научно-технической дисциплины — иссле-

дования операций. С методологической точки зрения исследование операций определяют [7] как научный подход к решению задач организационного управления в сложных человеко-машинных системах. При решении любой конкретной задачи применение методов исследования операций предполагает: 1) построение математических моделей для задач принятия решений и управления в сложных ситуациях или в условиях неопределенности; 2) изучение взаимосвязей, определяющих возможные последствия принимаемых решений, а также установление критериев эффективности, позволяющих оценивать относительное преимущество того или иного варианта действий.

Важным следствием применения методов исследования операций для решения широкого круга практических задач явилось выделение небольшого числа классов типовых задач. Вследствие частой повторяемости задач определенных классов были разработаны методы построения их моделей и получения решений на этих моделях.

В настоящее время различают задачи следующих классов: 1) распределения ресурсов; 2) управления запасами; 3) выбора маршрута; 4) построения расписаний; 5) массового обслуживания; 6) упорядочения и координации; 7) замены оборудования; 8) комбинаторные задачи и др.

Исследование построенных математических моделей совместно с экспериментальными исследованиями на действующих макетах системы позволяет обосновать оптимальную структуру системы, определить оптимальные значения ее параметров и убедиться в том, что выбранный вариант системы соответствует цели ее разработки и обладает требуемой эффективностью. Таким образом могут быть решены основные задачи при проектировании БСУ.

Решение задач при макропроектировании БСУ предъявляет определенные требования к специализации инженеров-системотехников. В [51] сформулированы некоторые личные качества идеального системотехника: 1) ярко выраженная склонность к системной точке зрения (не просто внешний интерес), т. е. склонность объять систему в целом — видеть ее внутренние связи и поведение, понимать цели и задачи, стоящие перед системой; 2) способность суждения, трезвой оценки, выносимой с полной объективностью; 3) богатство фантазии и воображения, так как системной работе присуще творчество; 4) коммуникабельность, которая выражается не только в способности «ладить» с людьми (такт, чуткость, дипломатичность), но и в проявлении

положительных свойств лидера; 5) дар выражения — устного, письменного, иногда графического.

В научном плане инженер-системотехник — специалист широкого профиля, обладающий достаточным кругозором в наиболее важных областях современной техники и высокой математической подготовкой. Способность моделировать еще не построенную систему и предсказывать таким путем с достаточной степенью точности будущие характеристики этой системы — наиболее характерные черты современного системотехника.

1-3. ПРИНЦИПЫ ПОСТРОЕНИЯ ОБОБЩЕННЫХ КРИТЕРИЕВ

Сложность и комплексность проблем, возникающих при проектировании больших систем автоматизированного управления, привели к тому, что вопросы формирования критериев для анализа и синтеза систем перестали быть только искусством, основанным на инженерной интуиции, а превратились в серьезное научное направление, важность которого возрастает с каждым днем [18].

На этапе проектирования БСУ трудности по построению обобщенных критериев, в первую очередь, возникают из-за того, что существует большое число критериев, не всегда согласованных между собой и подчас противоречивых в силу многостороннего характера поведения БСУ и множества ее свойств, интересующих при решении различных системотехнических задач. Выбранный критерий или система критериев является основой для принятия решений о выборе структуры, алгоритмов БСУ из некоторого множества, которое в дальнейшем будем называть множеством альтернатив. Общая постановка задачи принятия решений выглядит следующим образом:

А. Имеется некоторое множество альтернатив (вариантов построения системы) A , причем каждая альтернатива a характеризуется определенной совокупностью свойств $a_1, a_2, \dots, a_n$.

Б. Имеется совокупность критериев $q = (q_1, q_2, \dots, q_i, \dots, q_n)$, отражающих количественно множество свойств системы, т. е. каждая альтернатива характеризуется вектором $q(a) = [q_1(a), q_2(a), \dots, q_i(a), \dots, q_n(a)]$.

В. Необходимо принять решение о выборе одной из альтернатив (варианта), причем решение называется простым, если выбор производится по одному критерию, и

сложным, если выбранная альтернатива не является наилучшей по какому-то одному критерию, но может оказаться наиболее приемлемой для всей их совокупности.

Г. Задача принятия решения по выбору альтернативы на множестве критериев формально сводится к отысканию отображения φ , которое каждому вектору q ставит в соответствие действительное число

$$E = \varphi(q) = \varphi(q_1, q_2, \dots, q_n),$$

определяющее степень предпочтительности данного решения.

Оператор φ называют интегральным (обобщенным) критерием. Интегральный критерий присваивает каждому решению по выбору альтернативы соответствующее значение эффективности E . Это позволяет упорядочить множество решений по степени предпочтительности. Рассмотрим основные методы формирования обобщенных критериев.

1. Наиболее простой метод построения интегрального критерия заключается в том, что один из критериев q_k принимается в качестве обобщенного, а все остальные учитываются в виде ограничений, определяющих область допустимых альтернатив:

$$E = q_k; \quad q_i \geq q_i^{(0)}, \quad i = 1, 2, \dots, l; \\ q_i \leq q_i^{(0)}, \quad i = l+1, l+2, \dots, n; \quad i \neq k,$$

где $q^{(0)} = (q_1^{(0)}, q_2^{(0)}, \dots, q_n^{(0)})$ — вектор, определяющий допустимые значения по всем критериям.

В этом случае задача сравнения альтернатив по векторному критерию эффективности сводится к задаче принятия решений со скалярным критерием, а все остальные критерии переводятся в разряд ограничений. Альтернативы, не укладывающиеся в заданные границы, сразу же отбрасываются как неконкурентноспособные. Полученные практические рекомендации, очевидно, будут зависеть от того, как будут выбраны ограничения для вспомогательных критериев. В такой формулировке задача принятия оптимального решения при выборе альтернативы формулируется как задача математического программирования:

$$\max_{a \in A} [q_k(a)], \quad \text{либо} \quad \min_{a \in A} [q_k(a)] \quad \text{при} \\ q_i(a) \geq q_i^{(0)}, \quad i = 1, 2, \dots, l; \\ q_i(a) \leq q_i^{(0)}, \quad i = l+1, l+2, \dots, n; \quad i \neq k.$$

В зависимости от вида функций $q_k(a)$, $q_i(a)$ и множества A для решения задачи выбора оптимальной альтернативы используются методы линейного, линейного целочисленного, нелинейного, дискретного, динамического программирования, подробно рассматриваемые в исследовании операций [21].

В соответствии с данным подходом: а) выбор варианта построения вычислительного комплекса в автоматизированных информационных системах управления обслуживанием типа «Сирена» может потребовать, чтобы время ответа на запрос пользователя было минимальным; коэффициент, характеризующий простои комплекса, был не выше заданного; суммарная стоимость на его создание и эксплуатацию — не выше заданной; б) выбор варианта построения большой автоматизированной информационно-измерительной системы, решающей задачу контроля сложного оборудования, может потребовать того, чтобы критерий достоверности результатов контроля был максимальным при условии, что быстродействие системы контроля не ниже заданного, а суммарная стоимость на создание и эксплуатацию системы — не выше заданной; в) выбор плана работы промышленного предприятия в автоматизированных системах типа АСУП может потребовать, чтобы прибыль была максимальной, план по ассортименту выполнен, а себестоимость выпускаемой продукции — не выше заданной и т. д.

Основным недостатком предложенного подхода является то, что альтернативы оцениваются только по одному критерию, а значения других критериев, если они удовлетворяют ограничениям, не учитываются. Достоинство — в сравнительной простоте построения критерия.

2. В ряде случаев обобщенный показатель эффективности строят на основе использования аддитивных и мультипликативных преобразований над выбранной системой частных критериев q_i . В случае использования аддитивных преобразований

$$E = \varphi(q_1, q_2, \dots, q_n) = \sum_{i=1}^n b_i q_i,$$

где $b_1, b_2, \dots, b_n$ — положительные или отрицательные коэффициенты, причем положительные ставятся при тех критериях, которые желательно максимизировать, а отрицательные — при тех, которые желательно минимизировать, при условии, что ищется $E_{\max}$.

Аддитивное преобразование для построения интегрального критерия эффективности очень часто используется, если объединение различных частных критериев возможно на экономической основе и сравнение альтернатив производится в основном по экономическому критерию.

В случае использования мультипликативного преобразования обобщенный критерий формируется следующим образом:

$$E = \varphi(q_1, q_2, \dots, q_n) = \prod_{i=1}^n q_i^{\lambda_i},$$

где λ_i — некоторые вещественные числа.

Рассмотрим пример. Обобщенный критерий технико-экономической эффективности цифровых вычислительных машин, предложенный в [18], может быть построен следующим образом. Пусть $q_1 = T$, $q_2 = n(T)$, $q_3 = s(T)$, где T — время живучести ЭВМ, т. е. ее полное рабочее время; $n(T)$ — объем работы, выполненный ЭВМ за время ее живучести, в пересчете на число операций типового набора; $s(T)$ — суммарные затраты на изготовление, амортизацию и эксплуатацию ЭВМ за время T . Тогда обобщенный критерий эффективности ЭВМ, учитывающий и техническую и экономическую эффективность, может быть представлен в виде

$$E = \varphi(q_1, q_2, q_3) = q_3/(q_1 q_2 q_1^{-1}),$$

где $q_2 q_1^{-1}$ отражает эффективное быстродействие ЭВМ, причем является обобщенным критерием оценки технической эффективности ЭВМ и зависит от ее важнейших технических характеристик — системы операций, номинального быстродействия, емкости памяти, надежности, скорости работы устройств ввода-вывода.

Недостатком аддитивных и мультипликативных преобразований является то, что существует неограниченная возможность компенсации. Для ее уменьшения вводятся ограничения, определяющие наименьшие (наибольшие) допустимые значения частных критериев оптимальности, и поиск оптимальной альтернативы осуществляется на множестве альтернатив, удовлетворяющих этим ограничениям. В случае, если критерии q_i принимают только значения 0; 1, их объединение может быть произведено по аналогии с аддитивным и мультипликативным преобразованиями на основе логических преобразований типов дизъюнкции и конъюнкции.

3. В некоторых случаях построение обобщенного критерия основано на том, что обобщенное качество альтернатив оценивается расстоянием между идеальной и рассматриваемой альтернативами и чем ближе качество рассматриваемой альтернативы к идеальной, тем она лучше. В качестве идеальной обычно принимается альтернатива, которой соответствует вектор $q = (q_1^{(0)}, q_2^{(0)}, \dots, q_n^{(0)})$, где компонентами являются максимальные значения для максимизируемых и минимальные значения для минимизируемых критериев оптимальности, достижимые на множестве альтернатив A с учетом современного уровня техники или с учетом прогресса. В этом случае обобщенные критерии могут быть сформулированы в виде:

а) суммы абсолютных отклонений от идеальной альтернативы для частных критериев одной размерности

$$E = \varphi(q_1, q_2, \dots, q_n) = \sum_{i=1}^l (q_i^{(0)} - q_i) + \sum_{i=l+1}^n (q_i - q_i^{(0)}), \quad (1-1)$$

где $q_i (i = 1, 2, \dots, l)$ — частные критерии оптимальности, подлежащие максимизации; $q_i (i = l+1, l+2, \dots, n)$ — частные критерии оптимальности, подлежащие минимизации;

б) суммы относительных отклонений для частных критериев различной размерности

$$E = \varphi(q_1, q_2, \dots, q_n) = \sum_{i=1}^l \frac{q_i^{(0)} - q_i}{q_i^{(0)} - q_i^{\min}} + \sum_{i=l+1}^n \frac{q_i - q_i^{(0)}}{q_i^{\max} - q_i^{(0)}}, \quad (1-2)$$

где $q_i^{\min}$, $q_i^{\max}$ — наименьшие значения для максимизируемых и наибольшие для минимизируемых критериев оптимальности по всему множеству альтернатив;

в) наибольшего абсолютного отклонения от идеального для частных критериев одной размерности

$$E = \varphi(q_1, q_2, \dots, q_n) = \max_i |q_i^{(0)} - q_i|; \quad (1-3)$$

г) наибольшего относительного отклонения от идеального для частных критериев различной размерности

$$E = \varphi(q_1, q_2, \dots, q_n) = \max_{i, j} \left(\frac{q_i^{(0)} - q_i}{q_i^{(0)} - q_i^{\min}}; \frac{q_j - q_j^{(0)}}{q_j^{\max} - q_j^{(0)}} \right), \quad (1-4)$$

$$i = 1, 2, \dots, l; j = l+1, l+2, \dots, n.$$

Перечисленные выше критерии (1-1) и (1-2) также обладают неограниченной возможностью компенсации, а критерии (1-3), (1-4) на дискретных множествах альтернатив могут дать неверные результаты выбора.

Рассмотренные выше в пунктах 1, 2, 3 основные способы построения интегральных критериев на основе формальных правил не учитывали ценности, полезности частных критериев q_i , используемых при решении задачи выбора альтернативы.

4. При построении обобщенного показателя эффективности в соответствии с теорией полезности объединение критериев q_i производится чаще всего на основе аддитивного преобразования

$$E = \varphi(q_1, q_2, \dots, q_n) = \sum_{i=1}^n b_i q_i.$$

Однако в этом случае значения коэффициентов b_i отражают полезность (ценность) критерия q_i при принятии сложного решения о выборе альтернативы. Определение их значений производится в результате предварительного опроса группы из m экспертов (специалистов в данной области). Один из возможных путей получения этих значений заключается в следующем. Каждый j -й эксперт вначале определяет набор чисел C_{ij} , отражающих его мнение об относительной ценности i -го критерия, причем числа C_{ij} записаны в произвольном масштабе. Затем они масштабируются, в результате получаем

$$b_{ij} = C_{ij} / \sum_{i=1}^n C_{ij}; \quad \sum_{i=1}^n b_{ij} = 1. \quad (1-5)$$

Окончательные значения коэффициентов b_i вычисляются в результате осреднения значений b_{ij} ($j = 1, 2, \dots, m$), получаемых от всех экспертов. Если компетентность экспертов в группе считается одинаковой, то

$$b_i = \frac{1}{m} \sum_{j=1}^m b_{ij}. \quad (1-6)$$

Если же компетентность j -го эксперта оценивается числом g_j , $\sum_{j=1}^m g_j = 1$, то

$$b_i = \sum_{j=1}^m g_j b_{ij}.$$

Рассмотрим основные методы формирования коэффициентов C_{ij} , отражающих мнение j -го эксперта о ценности i -го критерия. В дальнейшем предполагается, что вначале каждый эксперт провел ранжировку всех критериев, т. е. упорядочил их в соответствии с относительной ценностью так, что на первом месте находится самый главный критерий.

Метод ранжировки. В соответствии с данным методом производится нумерация всех критериев полученного ряда, причем все неразличимые критерии, которые оказались на одном месте, нумеруются в произвольном порядке. В результате данной процедуры каждый критерий получает свой номер. Ранг критерия определяется его номером, если на его месте в ряду отсутствуют какие-либо другие. Если на одном месте находится несколько неразличимых критериев, то ранг каждого из них равен среднему арифметическому их новых номеров.

Рассмотрим пример. Пусть имеем следующий ряд упорядоченных критериев $q_1, q_2, \dots, q_8$ для j -го эксперта:

$$q_3; \left| \begin{array}{c} q_5 \\ q_6 \end{array} \right|; \left| \begin{array}{c} q_2 \\ q_4 \end{array} \right|; q_8; q_7; q_1.$$

Ранги критериев, вычисленные в соответствии с вышеуказанной процедурой, сведены в табл. 1-1.

Таблица 1-1

i	1	2	3	4	5	6	7	8
r_{ij}	8,0	4,5	1,0	4,5	2,5	2,5	7,0	6,0

Переход от рангов к коэффициентам C_{ij} производится на основе гипотезы о линейной зависимости между рангом и относительной ценностью критерия. Чем ниже ранг, тем более важным является соответствующий критерий. Определение коэффициентов C_{ij} для произвольного r_{ij} ($1 \leq r_{ij} \leq h$) производится в соответствии со следующей формулой:

$$C_{ij} = 1 - \frac{r_{ij} - 1}{n}.$$

Для рассмотренного примера коэффициенты C_{ij} сведены в табл. 1-2.

Таблица 1-2

i	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8
C_{ij}	0,125	0,433	1,000	0,433	0,812	0,812	0,250	0,375

Следует отметить, что гипотеза о линейной зависимости между рангом и относительной ценностью критерия делает оценки C_{ij} весьма грубыми, но определяет их сравнительно высокую достоверность.

Метод непосредственной оценки. В основу этого метода положена менее жесткая гипотеза об убывающей (но необязательно линейной) зависимости между рангом и относительной ценностью критерия. Вначале каждый j -й эксперт производит упорядочение всех критериев в соответствии с тем, что рассмотренной процедурой. После этого он эвристическим путем дает численную оценку относительной полезности каждого критерия по сравнению с самым главным, которому присваивается значение, равное единице. Всем неразличимым критериям присваиваются одинаковые значения C_{ij} . В результате каждому критерию в упорядоченном ряду вместо рангов сразу присваиваются числа C_{ij} , совокупность которых должна образовать невозрастающую последовательность. При использовании метода непосредственных оценок возникает возможность более дифференцированно подходить к оценке важности отдельных критериев, но при этом понижается достоверность полученной информации.

Метод последовательных предпочтений. Алгоритм последовательных предпочтений предназначен для повышения достоверности информации, полученной от экспертов методом непосредственных оценок. Он позволяет каждому эксперту провести самоконтроль суждений на основе сопоставления трех подходов: ранжирования критериев, числовой оценки их ценности и сравнения $n - 2$ пар специально подобранных абстрактных объектов.

Последняя процедура, отражающая сущность метода последовательных предпочтений, основана на следующей гипотезе [1]. Если ценность i -го критерия объекта некоторого класса для j -го эксперта есть C_{ij} , то ценность объекта

по всем критериям определяется $\sum_{i=1}^n C_{ij}$. В процессе коррекции оценок эксперт должен ответить на ряд почти дихотомических вопросов: для $i = 1, 2, \dots, (n - 2)$ какой из двух объектов лучше — обладающий только i -м критерием или совокупностью из $(i + 1, i \pm 2, \dots, n)$ критериев? В зависимости от ответа на i -й вопрос составляется одно из трех соотношений:

$$C_{ij}R \sum_{k=i+1}^n C_{kj}, \text{ где } R \in [>, <, =].$$

В результате будут получены $(n - 2)$ условия:

$$C_{1j}R \sum_{k=2}^n C_{kj}; C_{2j}R \sum_{k=3}^n C_{kj}, \dots, C_{n-2,j}R (C_{n-1,j} + C_{n,j}).$$

Далее производится последовательная проверка каждого из этих условий, начиная с последнего, на соответствие ранее выбранным оценкам C_{ij} и их ранжировке. При выявлении противоречий в i -м условии эксперт должен либо изменить знак отношения R , либо откорректировать значение величины C_{ij} . В последнем случае он обязан убедиться в том, что не оказалась нарушенной первоначальная ранжировка критериев. При нарушении ее необходимо либо изменить порядок критериев, либо откорректировать значение C_{ij} . После исправления последней оценки C_{ij} ее значение может отличаться от единицы. Следует отметить, что в этом случае психологические ограничения не дают использовать метод последовательных предпочтений, когда число рассматриваемых критериев превышает семь. Рассмотрим пример.

Пример 1-1. Пусть некоторый эксперт выставил следующий ряд коэффициентов C_i , отражающих его мнение об относительной ценности шести частных критериев некоторого объекта (табл. 1-3).

Таблица 1-3

i	1	2	3	4	5	6
C_i	1,0	0,9	0,7	0,6	0,3	0,1

Для уточнения оценок коэффициентов C_i эксперту предлагается сравнивать четыре пары абстрактных объектов. Каждому объекту

соответствует вектор $x = (x_1, x_2, \dots, x_i, \dots, x_6)$, где $x_i = (0; 1)$: 1 — учитывается полезность i -го критерия, 0 — не учитывается; тогда имеем:

- 1) (100000) хуже (011111);
- 2) (010000) лучше (001111);
- 3) (001000) хуже (000111);
- 4) (000100) лучше (000011).

Эксперт вынес систему решений. Соотношение $x^{(1)}$ лучше $x^{(2)}$ соответствует большей предпочтительности для эксперта объекта $x^{(1)}$ по сравнению с объектом $x^{(2)}$.

Непротиворечивость принятых решений должна подтверждаться выполнением системы неравенств:

$$1) C_1 < \sum_{k=2}^6 C_k; 2) C_2 > \sum_{k=3}^6 C_k; 3) C_3 < \sum_{k=4}^6 C_k; 4) C_4 > C_5 + C_6.$$

Проверку неравенств начинаем с последнего (четвертого). Третье и четвертое неравенства выполняются, второе — нет; значит, необходимо скорректировать значения коэффициента C_2 . Примем значение $C_2 = 2$. Однако одновременно необходимо изменить значение C_1 таким образом, чтобы, во-первых, сохранился первоначальный порядок критериев, определенный экспертом, т. е. $C_1 > C_2$, и, во-вторых, выполнялось первое неравенство. Принимаем, например, значение $C_1 = 2,5$. В результате применения метода последовательных предпочтений получили непротиворечивый ряд оценок (табл. 1-4), которые в дальнейшем необходимо масштабировать.

Таблица 1-4

i	1	2	3	4	5	6
C_i	2,5	2,0	0,7	0,6	0,3	1,0

При использовании всех рассмотренных выше методов возникает естественный вопрос: насколько можно доверять результатам оценки коэффициентов C_{ij} , полученным из субъективных мнений экспертов? Достоверность результатов экспертного анализа чаще всего характеризуется степенью согласованности данных ими оценок. Для количественной оценки степени согласованности часто используется коэффициент конкордации

$$W = \frac{12S}{m^2(n^3 - n)},$$

где $s = \sum_{i=1}^n \left[\sum_{j=1}^m r_{ij} - 0,5m(n+1) \right]^2$; r_{ij} — место, которое заняло i -е свойство в ранжировке j -м экспертом.

Коэффициент W позволяет оценить, насколько согласованы между собой ряды предпочтительности, построенные каждым экспертом. Его значение находится в пределах $0 \leq W \leq 1$, причем $W = 0$ означает полную противоположность, а $W = 1$ — полное совпадение ранжировок. Практическая достоверность считается хорошей, если $W = 0,7 \div 0,8$.

На основе рассмотренных методов могут быть определены значения коэффициентов C_{ij} ($j = 1, 2, \dots, m, i = 1, 2, \dots, n$), по которым в соответствии с формулами (1-5), (1-6) будут вычислены коэффициенты b_i линейной формы интегрального критерия. При использовании такого подхода к формированию интегрального критерия будем в дальнейшем считать, что единица измерения каждого свойства системы, отраженного в соответствующем частном критерии, выбрана по принципу «чем больше, тем лучше». Отсюда будет следовать, что качество решения по выбору альтернативы тем лучше, чем больше значение показателя эффективности.

Так как критерии q_i могут иметь различную размерность, то при использовании их в качестве аргументов функции E необходимо провести нормирование, т. е. привести их к общей размерности, и в частности к безразмерному виду.

Для придания равномерности влияния каждого из критериев на значение интегрального критерия необходимо выравнивать диапазоны изменения значений критериев путем масштабирования и сведения их к диапазону $[0; 1]$.

Проведение преобразований типов нормирования и масштабирования требует, чтобы для каждого из критериев были определены понятия «негодного» и «идеального» объектов, а это означает, что должны быть заданы допустимые области изменений значений критериев q_i , $q_{in} < q_i \leq q_{ib}$. В этом случае самым простым масштабирующим и нормирующим преобразованием является линейное преобразование следующего вида:

$$q_i^{\text{отн}} = \begin{cases} 0 & \text{при } q_i \leq q_{in}; \\ \frac{q_i - q_{in}}{q_{ib} - q_{in}} & \text{при } q_{in} < q_i \leq q_{ib}; \\ 1 & \text{при } q_i > q_{ib}, \end{cases}$$

где $q_i^{\text{отн}}$, q_{in} , q_{ib} — относительное, нижнее и верхнее значения критерия q_i соответственно.

В случае такого преобразования чувствительность шкалы изменения q_i во всем диапазоне изменений q_i постоянна. Если же нас особенно интересуют альтернативы в окрестности некоторой точки q_i^* , то можно повысить разрешающую способность частного критерия в окрестности этой точки за счет использования соответствующих нелинейных преобразований (рис. 1-6).

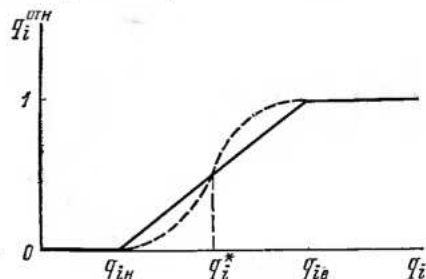


Рис. 1-6. Приведение к диапазону [0; 1] при использовании нелинейных преобразований

После проведения операций нормирования и масштабирования область годных альтернатив окажется заданной в виде n -мерного единичного куба, причем

$$0 < q_i^{\text{отн}} \leq 1, \quad i = 1, 2, \dots, n.$$

В результате проведенных преобразований для каждой рассматриваемой альтернативы будет определен вектор $q^{\text{отн}}(a)$, причем $a \in A$, где A — множество возможных альтернатив. Оценка интегрального показателя решения по выбору альтернативы производится в соответствии со следующим соотношением:

$$E = \varphi [q^{\text{отн}}(a)] = \sum_{i=1}^n b_i q_i^{\text{отн}}(a),$$

где b_i — коэффициенты, определяемые в соответствии с формулами (1-5), (1-6).

Заканчивая рассмотрение вопросов, связанных с построением обобщенного критерия эффективности сложных систем на основе метода экспертных оценок, следует еще раз обратить внимание на то, что на многих этапах его построения приходится опираться на субъективные мнения специалистов при выборе: наиболее существенных частных критериев; процедуры и единицы измерения для критериев;

«идеального» значения критерия, а также значения, дающего наибольшую разрешающую способность критерия; нормирующего и масштабирующего преобразований; структуры функции обобщенного критерия; значений весовых коэффициентов. Поэтому решения на каждом из перечисленных этапов должны приниматься на основе усредненного мнения многих специалистов, что повышает объективное содержание критерия. Однако это не исключает таких ситуаций, когда оценки некоторых реальных объектов, полученные с помощью обобщенного критерия, противоречат мнению специалистов. В подобных случаях следует не отказываться от дальнейшего использования данного подхода, а тщательно проанализировать и выявить конкретные причины расхождения, после чего внести соответствующие изменения в критерий.

Изложенные в данной главе вопросы являются основой для решения различных системотехнических задач, возникающих на этапе макропроектирования БСУ. В последующих главах рассматриваются задачи формализации описания структуры БСУ и методы определения их основных структурно-топологических характеристик, раскрывается содержание основных подходов к построению моделей функционирования БСУ для анализа различных функциональных свойств, исследуются различные постановки структурного синтеза.

ГЛАВА ВТОРАЯ

СТРУКТУРНЫЙ АНАЛИЗ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

2-1. ЦЕЛИ И ЗАДАЧИ СТРУКТУРНОГО АНАЛИЗА

Одной из важнейших характеристик всякой системы является ее структура. Под структурой системы, как уже было ранее отмечено, будем понимать совокупность элементов и связей между ними, которые определяются исходя из распределения функций и целей, поставленных перед системой. Примеры БСУ, приведенные в гл. 1, свидетельствуют о большом числе и разнообразии элементов и их связей. Поэтому одной из главных задач структурного

анализа является построение наглядной формальной модели, отображающей существующую систему отношений элементов как между собой, так и с внешней средой. Структурная модель БСУ чаще всего является многоуровневой, причем конкретизация структуры дается на стольких уровнях, сколько их требуется для создания полного представления об основных свойствах системы (рис. 2-1).

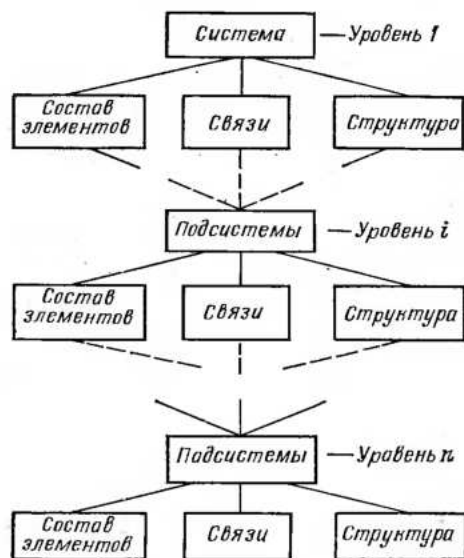


Рис. 2-1. Многоуровневое представление структуры БСУ

При создании АСУП структурная модель рассматривается на нескольких крупных уровнях: организации, функций управления, технических средств. В связи с этим вводятся понятия организационной, функциональной и технической структур.

При анализе организационной структуры предприятия как объекта управления решаются следующие задачи: 1) описание состава организации и построение ее структурной схемы; 2) определение функций отдельных подразделений, раскрытие их структурной схемы; 3) описание материальных, вещественных и информационных связей; 4) построение обобщенной структурной информационной модели предприятия.

При анализе функциональной структуры: 1) изучаются функции управления в структурных подразделениях существующей системы; 2) выбирается состав автоматизируемых функций; 3) определяются их взаимосвязи; 4) составляется обобщенная функциональная структура задач управления АСУП.

При анализе технической структуры: 1) определяются основные элементы, участвующие в основных информационных процессах: регистрации и подготовке информации, сборе и передаче, хранении и обработке, воспроизведении и выдаче информации; 2) составляется формальная структурная модель системы технических средств с учетом топологии расположения элементов системы и информационного и энергетического взаимодействия их как между собой, так и с внешней средой.

Независимо от уровня рассмотрения общая задача структурного анализа состоит в том, чтобы исходя из заданного описания элементов системы и непосредственных связей между ними получить заключение о структурных свойствах системы в целом и основных ее подсистем. При решении практических задач структурного анализа БСУ обычно принимаются три уровня описания связей между элементами: 1) наличие связи; 2) направление связи; 3) вид и направление сигналов, определяющих взаимодействие элементов.

На первом уровне, когда исходят лишь из наличия или отсутствия связей между элементами, изучаемой системе может соответствовать неориентированный граф, вершинами которого являются элементы системы, а ребрами — существующие непосредственные связи между элементами. Основные задачи структурного анализа на этом уровне сводятся к следующим: 1) определение связности (целостности) системы; если система не является связной, то ставят задачу выделения изолированных связных подсистем со списками входящих в них элементов; 2) выделение циклов; 3) определение минимальных и максимальных последовательностей элементов (цепей), разделяющих элементы друг от друга.

На втором уровне, когда задано направление связи, системе соответствует ориентированный граф, направления дуг которого совпадают с направлениями связей. На этом уровне результаты структурного анализа оказываются более содержательными. К задачам структурного анализа в этом случае относят: 1) определение связности

системы; 2) топологическую декомпозицию с выделением сильно связанных подсистем; 3) перечисление входных и выходных полюсов и в соответствии с этим выделение узлов приема и выдачи информации; 4) выделение уровней в структуре и определение их взаимосвязи; 5) определение максимальных и минимальных путей; 6) определение характеристик топологической значимости элементов; 7) получение информации о слабых местах структуры и др.

На третьем уровне описания связей между элементами системы учитывается не только направленность связи, но и раскрываются состав и характер сигналов взаимодействия элементов (входные, выходные и управляющие).

Указанные выше задачи структурного анализа первого и второго уровней расширяются за счет включения следующих: 1) выделение местных и общих контуров управления; 2) при многорежимном характере функционирования выделение типичных структурных конфигураций для каждого из режимов; 3) выделение путей непосредственной передачи входных сигналов и путей передачи управляющих сигналов и др.

Рассмотрим сначала более подробно задачи структурного анализа при минимальном объеме априорной информации о структуре системы, когда учитывается только факт наличия связи и ее направленность. Основы подобной формализации описания структур заложены в теории графов [28].

2-2. ФОРМАЛИЗАЦИЯ ОПИСАНИЯ СТРУКТУРЫ НА ОСНОВЕ ТЕОРИИ ГРАФОВ

Принцип представления структуры в виде графа весьма прост. Элементам системы при решении одних задач ставят в соответствие ребра графа, а связям — вершины графа. При решении других задач поступают иначе: элементам ставят в соответствие вершины графа, а связям — ребра графа. Граф, полученный в первом случае, называется *реберным*, во втором случае, — *вершинным*.

Рассмотрим некоторые основные определения и понятия из теории графов, необходимые при решении задач структурного анализа БСУ.

Определение графа, виды графов. Пусть определено некоторое множество элементов V . Граф $G = G(V)$ считается определенным, если задано некоторое семейство сочетаний элементов или пар вида $E = (a, b)$, где $a, b \in V$, указы-

вающее, какие элементы считаются связанными. В соответствии с геометрической интерпретацией пара $E = (a, b)$ называется *ребром*, а элементы a, b — *концевыми точками* ребра или *вершинами*. Если порядок расположения концов безразличен, т. е. если $E = (a, b) = (b, a)$, то говорят, что E есть *неориентированное ребро*; если же этот порядок важен, то E называют *ориентированным ребром* — *дугой*; при этом a называют *начальной вершиной*, а b — *конечной вершиной*. В теории графов принята также следующая терминология: ребро E *инцидентно* вершинам a, b , а вершины a, b *инцидентны* ребру E . Граф, составленный только из неориентированных ребер, называется *неориентированным*, а граф, составленный только из ориентированных ребер, *ориентированным*. Графы, у которых часть ребер ориентированна, часть — неориентированна, называются *смешанными*. Неориентированный граф G может быть превращен в ориентированный при помощи процесса удвоения, состоящего в замене каждого ребра E парой ребер с теми же концами и приписывании им противоположных ориентаций. Граф называется *конечным*, если число ребер конечно, и *бесконечным* — в противном случае. Граф, состоящий из изолированной вершины, называется *нульграфом*, а граф, ребрами которого являются всевозможные пары для двух различных вершин a, b из V , называется *полным графом*. В ориентированном полном графе имеются пары ребер по одному в каждом направлении, соединяющие любые две различные вершины (a, b) .

Способы формализованного задания графа. А. Графическое представление. Данный способ является наиболее наглядной формой представления отношений между элементами, однако он не может быть использован при решении задач структурного анализа с использованием ЭВМ.

Б. Матричное представление. Существуют различные формы матричного представления графа $G = G(V)$. Матрица смежности вершин для неориентированного графа имеет вид $A = \|a_{ij}\|$, где n — число вершин в графе, а элементы a_{ij} определяются следующим образом:

$$a_{ij} = \begin{cases} 1 & \text{при наличии связи;} \\ 0 & \text{при отсутствии связи.} \end{cases}$$

При этом предполагается, что нумерация вершин графа уже проведена. Для неориентированного графа матрица

смежности является симметрической, что следует учитывать при использовании ЭВМ для компактного использования оперативной памяти в задачах структурного анализа.

В матрице смежности вершин для ориентированного графа $A = \|a_{ij}\|$ элементы определяются следующим образом:

$$a_{ij} = \begin{cases} 1, & \text{если из вершины } i \text{ можно перейти в вершину } j; \\ 0 & \text{в противном случае.} \end{cases}$$

Вид матрицы смежности ориентированных графов будет существенным образом зависеть от выбранного алгоритма нумерации вершин и, как показано в [28], выбрав определенный принцип нумерации вершин на графе для некоторых видов графов (графов без контуров), можно свести матрицу A к треугольному виду, где $a_{ij} = 0$ при $j > i$, что следует также учитывать при задании исходной информации о структуре системы для ЭВМ. Данный вопрос будет далее рассмотрен при анализе способа введения порядковых функций на графах.

Можно аналогичным образом построить соответствующие матрицы смежности ребер, где элемент a_{ij} считается ненулевым, если ребра графа имеют общую вершину.

Матрица инцидентности $B = \|b_{ij}\|$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$, где n — число вершин, а m — число ребер, определяется для неориентированного графа следующим образом:

$$b_{ij} = \begin{cases} 1, & \text{если } i\text{-я вершина инцидентна} \\ & \text{данному } j\text{-му ребру (есть связь);} \\ 0, & \text{если нет связи;} \end{cases}$$

а для ориентированного графа:

$$b_{ij} = \begin{cases} 1, & \text{если } i\text{-я вершина есть начало } j\text{-го ребра;} \\ -1, & \text{если } i\text{-я вершина есть конец } j\text{-го ребра;} \\ 0, & \text{если } i\text{-я вершина не инцидентна } j\text{-му ребру.} \end{cases}$$

Для многих практических задач, связанных с анализом структурных свойств БСУ, большинство элементов в матрицах A и B оказываются нулевыми, что приводит к весьма неэффективному использованию памяти ЭВМ при задании исходной информации о графе. В этом случае целесообразно использовать множественные способы формализации структуры.

В. Множественное представление. В этом случае для ориентированного графа $G(V)$ задается множество вершин V и соответствие G , которое показывает, как связаны между собой вершины. Соответствие G в этом случае называется отображением множества V в V . Для каждой вершины i соответствие G определяет множество вершин $G(i)$, в которые можно непосредственно попасть из вершины i . В ряде случаев $G(i)$ называется множеством правых инцидентий.

Множество $G^{-1}(i)$ определяет все вершины, из которых можно непосредственно попасть в вершину i , и поэтому называется обратным соответствием (отображением). По аналогии с $G(i)$ множество $G^{-1}(i)$ называется множеством левых инцидентий.

Таким образом, ориентированный граф задается перечислением (списком) либо множеств вида $G(i)$, либо множеств вида $G^{-1}(i)$ для всех вершин графа. Такой способ оказывается значительно более компактным и эффективным при задании исходной информации о структуре для решения задач анализа и синтеза, особенно для иерархических систем.

Рассмотрим различные способы формализации исходной информации о структуре на примере.

Пример 2-1. Представление структуры системы в виде графа (рис. 2-2, а).

Матрица смежности вершин A выглядит следующим образом:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Матрица инцидентий B имеет вид:

$$B = \begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 & -1 & 0 & 0 \end{pmatrix}.$$

Множественное задание структуры системы сводится к следующей системе множеств: $G(1) = (2, 3)$; $G(2) = \emptyset$; $G(3) = (5, 4)$; $G(4) = 2$; $G(5) = (1, 2)$; $G^{-1}(1) = 5$; $G^{-1}(2) = (1, 5, 4)$; $G^{-1}(3) = (1)$; $G^{-1}(4) = (3)$; $G^{-1}(5) = (3)$.

Определение частичного графа и подграфа. Граф H называется **частичным графом** графа G , если его множество вершин $V(H)$ содержится во множестве вершин V графа G и все ребра графа H являются ребрами графа G . Граф $G_1(D)$ называется **подграфом** графа $G(V)$, если множество вершин D содержится во множестве вершин V и ребра графа G_1 являются всеми ребрами графа G , концы которых лежат во множестве D . Если $D = V$, то подграф $G(D)$ совпадает с графом $G(V)$.

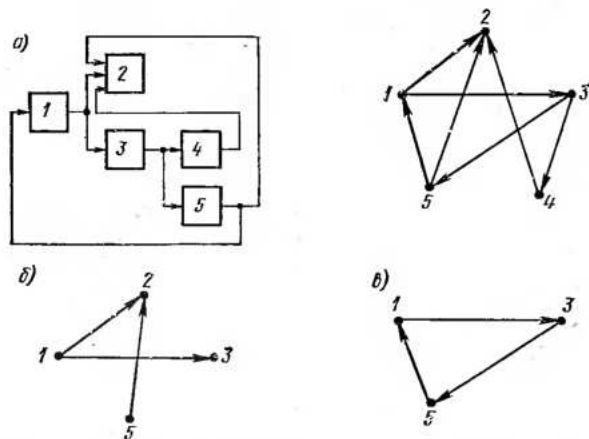


Рис. 2-2. Отображение структуры системы в виде ориентированного графа: а — структура и граф системы; б — частичный граф; в — подграф

Частичный граф H , выделенный из графа G (рис. 2-2, а), построен на рис. 2-2, б, а подграф G_1 — на рис. 2-2, в. Понятия частичного графа и подграфа являются основой для выделения подсистем при структурном анализе БСУ.

Определение цепи, пути, цикла, контура. Цепью называется такая последовательность ребер $E_0, E_1, \dots, E_{k-1}, E_k, \dots$, когда каждое ребро E_{k-1} соприкасается одним из концов с ребром E_k ; цепь можно обозначить последовательностью вершин, которые она содержит. Например, для графа (рис. 2-3) последовательности вершин $(1, 4, 3, 5)$, $(1, 3, 5)$ являются цепями. Понятие цепи обычно используется для неориентированных графов. Путем называется такая последовательность дуг, когда конец каждой предыдущей дуги совпадает с началом последующей;

например, для графа (рис. 2-2, а) последовательность дуг $(1, 3), (3, 5), (5, 1)$ является путем. Понятие пути обычно используется для ориентированных графов. Циклом называется такая конечная цепь, которая начинается и заканчивается в одной вершине; например, для графа (рис. 2-3) цепь $(1, 4, 3, 1)$ является циклом; данное понятие имеет смысл только для неориентированных графов. Контуром называется такой конечный путь, у которого начальная вершина первой дуги совпадает с конечной вершиной последней дуги пути; для графа (рис. 2-2, а) последовательность дуг $(1, 3), (3, 5), (5, 1)$ — контур. Длиной цепи (пути) называют число ребер (дуг), входящих в цепь (путь) графа.

Матрица смежности вершин A является матрицей непосредственных путей графа, имеющих длину, равную единице. Общее число транзитных путей от вершины i к вершине j длиной k может быть получено в результате возведения в k -ю степень матрицы A :

$$A^k = AA \dots A = A^{k-1}A. \quad (2-1)$$

Элемент матрицы $A^k a_{ij}^{(k)}$ определяет число путей длиной k от вершины i к вершине j .

Степень вершины. Число ребер, инцидентных вершине неориентированного графа, называют **степенью вершины** $\rho(i)$. Число дуг ориентированного графа, которые имеют своей начальной вершиной вершину i , называют **полустепенью исхода** вершины i и обозначают $\rho^+(i)$. Аналогично, число дуг, которые имеют своей конечной вершиной вершину j , называют **полустепенью захода** вершины j и обозначают $\rho^-(j)$. Совершенно очевидно, что сумма полустепеней захода всех вершин графа, а также сумма полустепеней исхода всех вершин равны общему числу дуг ориентированного графа $G(V)$, т. е.

$$\sum_{i=1}^n \rho^+(i) = \sum_{j=1}^n \rho^-(j) = m,$$

где m — число дуг графа, а n — число вершин графа.

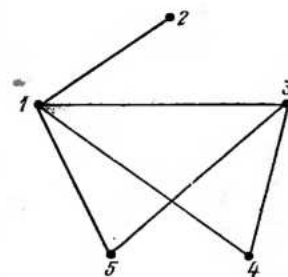


Рис. 2-3. Неориентированный граф

Для неориентированного графа имеем

$$\frac{1}{2} \sum_{i=1}^n \rho(i) = m. \quad (2-2)$$

Понятие связности графа. Для неориентированных графов вводится понятие слабой связности, или просто связности. Граф $G(V)$ называется **слабо связным** (связным), если для любых вершин графа i и j существует цепь из вершины i в вершину j . Для ориентированных графов вводится дополнительно еще понятие сильной связности.

Граф $G(V)$ называется **сильно связным**, если для любых вершин графа i, j существует путь из вершины i в вершину j . Граф, изображенный на рис. 2-3, является слабо связным; граф на рис. 2-4, а — сильно связным, а граф на рис. 2-4, б — несвязным, распадающимся на два сильно связных подграфа.

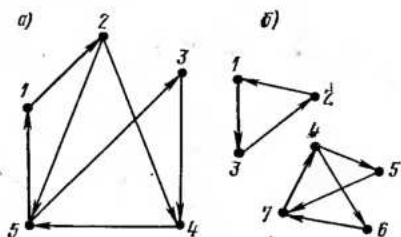


Рис. 2-4. Сильно связный граф (а) и несвязный граф, распадающийся на сильно связные подграфы (б)

совокупность связных подграфов; а любой ориентированный граф — на совокупность сильно связных подграфов.

Порядковая функция на графе. Понятие уровня. Целью введения порядковой функции на графе без контуров является разбиение множества вершин графа на непересекающиеся подмножества, упорядоченные так, что если вершина входит в подмножество с номером i , то следующая за ней вершина — в подмножество с номером, большим i . Полученные непересекающиеся подмножества называются **уровнями** u **р** **о** **в** **н** **я** **м** **и**. Алгоритм упорядочения сводится к следующему.

1. В подмножество нулевого уровня N_0 включаются все вершины i , у которых $G^{-1}(i) = \emptyset$ (пустое подмножество). Проводится последовательная нумерация вершин: $1, 2, \dots, l$.

2. В подмножество первого уровня N_1 включаются все вершины i , у которых $G^{-1}(i) \subset N_0$. Проводится последовательная нумерация вершин: $l+1, l+2, \dots, l+r$.

3. В подмножество второго уровня N_2 включаются все вершины i , у которых $G^{-1}(i) \subset (N_0 \cup N_1)$. Проводится последовательная нумерация вершин: $l+r+1, \dots, l+r+p$.

4. В подмножество третьего уровня N_3 включаются все вершины i , у которых $G^{-1}(i) \subset (N_0 \cup N_1 \cup N_2)$, после чего также проводится нумерация вершин.

Данный процесс повторяется до тех пор, пока не будут пронумерованы все вершины графа. Изложенная процедура нумерации приводит к тому, что в матрице смежности вершин графа $a_{ij} = 0$ при $i > j$.

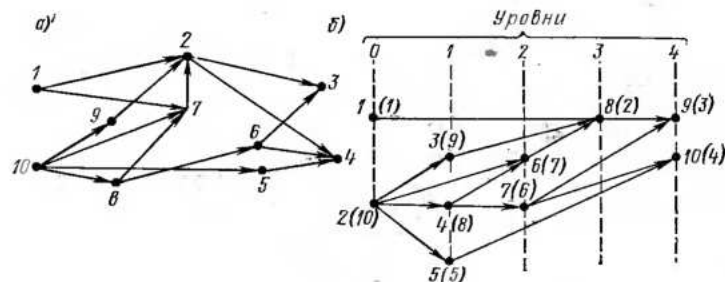


Рис. 2-5. Введение порядковых функций на графе: а — неупорядоченный граф; б — упорядоченный граф

На рис. 2-5 показаны упорядоченный и неупорядоченный графы без контуров. Для графа при наличии контуров в соответствии с рассмотренным выше алгоритмом выделяются сначала сильно связные подграфы, образующие классы, а далее, так как граф классов не имеет контуров, на нем можно произвести упорядочение и ввести понятие уровней.

Числовая функция на графе. Числовую функцию на графе задают обычно либо на вершинах, либо на дугах (ребрах) графа.

Числовая функция на вершинах графа считается заданной, если каждой i -й вершине a_i графа $G(V)$, $a_i \in V$, ставится в соответствие некоторое число $l_i = l(a_i)$ из некоторого множества L .

Числовая функция на дугах (ребрах) для ориентированного (неориентированного) графа считается заданной, если каждой дуге ($a_i a_j$) или ребру ставится в соответствие число $q = q(a_i a_j)$ из некоторого множества Q . В некоторых

случаях числовая функция на графе задается комбинированным способом как на вершинах, так и на дугах.

Значение функции на пути S через вершины $a_1, a_2, \dots, a_i, \dots, (a_i \in S)$ при задании числовой функции на вершинах графа определяется либо в соответствии с аддитивной формой

$$l_S = \sum_{a_i \in S} l(a_i), \quad (2-3)$$

либо в соответствии с мультипликативной формой

$$l_S = \prod_{a_i \in S} l(a_i). \quad (2-4)$$

Аналогичным образом определяется значение функции на пути через вершины $a_1, a_2, \dots, a_i, \dots$ при задании числовой функции на дугах (ребрах) графа:

$$q_S = \sum_{(a_i a_j) \in S} q(a_i a_j); \quad (2-5)$$

$$q_S = \prod_{(a_i a_j) \in S} q(a_i a_j). \quad (2-6)$$

В соответствии с данными определениями может быть поставлена задача нахождения путей через множество вершин (дуг), обладающих определенным свойством, с максимальным (минимальным) значением числовой функции. Такие пути называются **максимальными** и **минимальными**. Определение максимальных (минимальных) путей на графе чаще всего формализуется в виде задачи динамического программирования. Так, например, определение максимального пути на графе без контуров в соответствии с (2-5) реализуется на основе следующего функционального уравнения динамического программирования:

$$q_S^{\max}(a_1 a_j) = \max_{a_i \in G^{-1}(a_j)} [q_S^{\max}(a_1 a_i) + q(a_i a_j)], \quad (2-7)$$

где $q_S^{\max}(a_1 a_j)$ — максимальное значение функции на путях S из некоторой начальной вершины a_1 в вершину a_j [аналогично $q_S^{\max}(a_i, a_j)$]; $G^{-1}(a_j)$ — множество левых инцидентов для вершины a_j ; $q(a_i a_j)$ — значение функции на дуге $(a_i a_j)$.

При использовании данного функционального уравнения обычно предполагается, что все вершины в графе упорядочены. Для иллюстрации применения уравнения (2-7) рассмотрим пример.

Пример 2-2. Определение максимального пути на графе.

Для графа (рис. 2-6) найти максимальный путь из вершины a_1 в вершину a_7 .

Для вершины a_1 принимаем $q_S^{\max}(a_1 a_1) = 0$. Для вершин a_2, a_3, a_4 : $q_S^{\max}(a_1 a_2) = 2$; $q_S^{\max}(a_1 a_3) = 3$; $q_S^{\max}(a_1 a_4) = 4$. Для вершины a_5 : $q_S^{\max}(a_1 a_5) = \max(2 + 1; 3 + 2; 4 + 1) = 5$. Для вершины a_6 : $q_S^{\max}(a_1 a_6) = \max(3 + 3; 4 + 4) = 8$. Для вершины a_7 : $q_S^{\max}(a_1 a_7) = \max(5 + 2; 8 + 1) = 9$.

Значение функции на максимальном пути равно девяти, а сам путь выделен жирной линией на рис. 2-6.

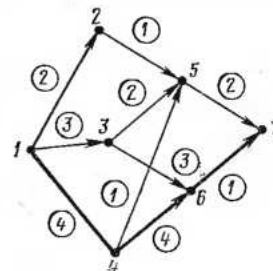


Рис. 2-6. Определение максимального пути на графе

Определение максимальных и минимальных путей имеет многочисленные приложения при проектировании БСУ: в задачах сетевого и календарного планирования для определения критического пути, в транспортных задачах, в задачах контроля и технической диагностики и др. Рассмотрим далее применение изложенных выше понятий теории графов для описания и анализа потоков информации в БСУ.

2-3. ОПИСАНИЕ И АНАЛИЗ ПОТОКОВ ИНФОРМАЦИИ В БСУ

Как уже было отмечено в гл. 1, большие автоматизированные системы управления, в первую очередь, отличаются информационной сложностью, т. е. для принятия решений требуется сбор, обработка и анализ больших объемов информации. В связи с этим большую актуальность приобретают задачи описания и анализа структуры потоков информации в системе.

В дальнейшем для большей конкретности будем рассматривать данный круг вопросов на примере АСУП. Источником информации в системе является документ. Взаимодействие элементов в системе приводит к тому, что одни документы формируются на основании других, т. е. происходит движение информации с целью получения

определенного функционального результата. Как для системы в целом, так и для любой подсистемы все документы могут быть классифицированы на: 1) исходные — поступающие в систему; 2) внешние — результаты переработки исходных; 3) промежуточные — результаты переработки исходных, которые используются для вычисления внешних документов, но сами из системы не выдаются. Совокупность исходных и внешних документов составляет информационный базис системы, который не зависит от программ обработки информации, а определяется функциями системы (подсистемы). Между документами, входящими в поток, существуют отношения вхождения и порядка.

Отношение вхождения означает $x_j = x_{j1}x_{j2} \dots x_{jn}$, т. е. документ x_j образуется непосредственно из документов $x_{j1}, x_{j2}, \dots, x_{jn}$, а отношение порядка: x_j следует за x_i — означает, что документ x_j может быть образован только после документа x_i . Если документам сопоставить вершины графа, а дугам — соответствующие отношения вхождения и порядка, то получится структура, отражающая информационное взаимодействие элементов системы, называемая и н ф о р м а ц и о н н ы м г р а ф о м.

Введение порядковой функции, как уже было показано в § 2-2, на информационном графе позволяет выявить его многоуровневую структуру и классифицировать документы по уровням формирования. Будем считать в дальнейшем, что структура информационного графа задана матрицей смежности A .

Формальный анализ свойств последовательности матриц $A, A^1, \dots, A^k, \dots, A^n$ позволяет выявить следующие свойства и параметры построенного информационного графа (n — число вершин).

1. Порядком π_j элемента j называется длина наибольшего пути, связывающего j -й элемент с i -м ($i = 1, 2, \dots, n$; $j \neq i$). Порядок π_j формально определяется по условиям: $\sigma^j(k = \pi_j) > 0$, $\sigma^j(k = \pi_j + 1) = 0$, где $\sigma^j(k)$ — сумма элементов j -го столбца матрицы A^k .

Физический смысл параметра π_j — номер такта, к которому «готовы» все документы, формирующие j -й документ.

2. Число $N = \max_j \pi_j$ называется порядком и н ф о р м а ц и о н н о г о г р а ф а. Если для N справедливы соотношения $A^N \neq 0$, $A^{N+1} = 0$, то соответствующий граф называется N -тактным.

3. Признаком контура служит появление ненулевых элементов на главной диагонали любой из матриц A^k ; вообще информационные графы чаще всего являются графами без контуров, поэтому наличие контура является формальным признаком ошибки в построении модели.

4. Равенство нулю суммы элементов j -го столбца матрицы смежности $\sigma^j(k = 1) = 0$ служит признаком для формального выделения исходных элементов, а значение $\sigma^j(k = 1)$, отличное от нуля, определяет число элементов, входящих в j -й элемент.

5. Аналогично четвертому свойству, равенство нулю суммы элементов i -й строки матрицы смежности информационного графа $\sigma_i(k = 1) = 0$ служит признаком для выделения функциональных (конечных) результатов, а значение $\sigma_i(k = 1)$, большее нуля, определяет число элементов, в которые входит i -й элемент.

6. Если при некотором $i = j$ одновременно $\sigma_i(k = 1) = 0$ и $\sigma^j(k = 1) = 0$, то к рассматриваемой схеме потока информации этот элемент не имеет отношения.

7. Число путей длиной k от элемента i к элементу j определяется элементом a_{ij}^k матрицы A^k .

8. Число всевозможных путей от элемента i к элементу j определяется элементом a_{ij}^Σ матрицы $A_\Sigma = \sum_{k=1}^N A^k$.

9. Отличные от нуля элементы j -го столбца матрицы A_Σ указывают все документы, участвующие в формировании документа j , а ненулевые элементы i -й строки матрицы A_Σ указывают все документы, при формировании которых используется документ i .

10. Максимальное значение порядка элементов i -й строки матрицы A , отличных от нуля, определяет номер такта τ_i , после которого элемент i уже не используется.

11. Число тактов, в течение которых элемент i должен храниться в памяти, равно: $t_i = \tau_i - \pi_i$.

12. Анализ структуры всех путей, существующих в информационном графе от i -го элемента к j -му, позволяет выявить как дублирующие связи, так и избыточные элементы.

Выявление прямого дублирования элементов и потенциально дублирующих связей необходимо для сокращения существующего числа элементов потока информации и повышения оперативности представления информации руководителям разных уровней управления.

Изложенные выше теоретические основы формализованного метода описания и анализа потоков информации позволили разработать инженерную методику его реализации с использованием ЭВМ, необходимую при построении АСУП [44].

2.4. ТОПОЛОГИЧЕСКАЯ ДЕКОМПОЗИЦИЯ СТРУКТУР

Проведение топологической декомпозиции структуры БСУ на основе представления ее в виде ориентированного графа $G(V)$ связано с выделением в ней отдельных сильно связанных подсистем. Для рассмотрения основного содержания алгоритма декомпозиции целесообразно ввести следующие понятия. Множество вершин, достижимых из вершины i , называется достижимым множеством $R(i)$ и определяется следующим образом:

$$R(i) = (i) \cup G(i) \cup G^2(i) \cup \dots \cup G^p(i) \cup \dots, \quad (2-8)$$

где $G(i)$ — множество вершин, достижимых из вершины i с использованием путей длиной, равной единице; $G^p(i)$ — множество вершин, достижимых из вершины i с использованием путей длиной p . При этом предполагается, что сама вершина i достижима с использованием пути длиной, равной нулю, и включена во множество $R(i)$.

В соответствии с (2-8) множество $R(i)$ может быть получено последовательным выполнением (слева направо) операций объединения множеств до тех пор, пока текущее множество $R(i)$ не перестанет увеличиваться по размеру при очередной операции объединения. С этого момента последующие операции не будут давать новых элементов множеству i , таким образом, будет образовано достижимое множество $R(i)$. Число объединений, конечно, зависит от графа, но очевидно, что число p не может быть больше числа вершин в графе.

Контрадостижимым множеством $Q(i)$ графа $G(V)$ называется множество таких вершин, когда из любой вершины этого множества можно достигнуть вершины i . Аналогично построению достижимого множества $R(i)$ можно построить множество $Q(i)$, используя следующее выражение:

$$Q(i) = (i) \cup G^{-1}(i) \cup G^{-2}(i) \cup \dots \cup G^{-p}(i) \cup \dots, \quad (2-9)$$

где $G^{-2}(i) = G^{-1}[G^{-1}(i)]$ — множество вершин, из которых можно достигнуть i -й вершины, при этом длина пути равна двум.

Операции выполняются слева направо до тех пор, пока очередная операция объединения не перестанет изменять текущее множество $Q(i)$.

Так как $R(i)$ является множеством вершин, достижимых из i -й вершины, а $Q(j)$ — множеством вершин, из которых можно достигнуть j -ю вершину, то множество $R(i) \cap Q(j)$ является множеством таких вершин, каждая из которых принадлежит по крайней мере одному пути, идущему от i -й вершины к j -й. Эти вершины называются существенными, или неотъемлемыми, относительно двух концевых вершин i и j . В свою очередь, множество $R(i) \cap Q(i)$ однозначно определяет сильно связный подграф графа $G(V)$, содержащий i -ю вершину, так как все существенные вершины, принадлежащие множеству $R(i) \cap Q(i)$, достижимы из i -й вершины i , кроме того, из каждой такой вершины достижима вершина i . Следовательно, все они взаимно достижимы. Из введенных выше определений вытекает основное содержание алгоритма декомпозиции.

В дальнейшем предполагаем, что нумерация вершин в графе произведена. Для первой вершины определяем множества $R(1)$ и $Q(1)$. Далее находим сильно связный подграф G_1 , включающий множество вершин $V_1 = R(1) \cap Q(1)$. Все вершины, принадлежащие $G_1(V_1)$, удаляются из первоначального графа $G(V)$. Далее эта процедура повторяется для новой вершины из оставшихся с меньшим номером. Эту процедуру необходимо повторять до тех пор, пока все вершины начального графа не будут сгруппированы в соответствующие сильно связанные подграфы. После завершения данного алгоритма граф $G(V)$ будет разбит на сильно связанные подграфы. Рассмотрим пример (рис. 2-7).

Пример 2-3. Найдем сильно связный подграф для вершины 1. Из (2-8) и (2-9) получаем: $R(1) = (1, 2, 4, 5, 6, 7, 8, 9, 10)$; $Q(1) = (1, 2, 3, 5, 6)$. Отсюда находим множество вершин для сильно связного подграфа, содержащего вершину 1: $V_1 = (1, 2, 5, 6)$.

Аналогичным образом находим остальные связанные подграфы: $V_2 = (3)$; $V_3 = (4, 7, 9)$; $V_4 = (8, 10)$.

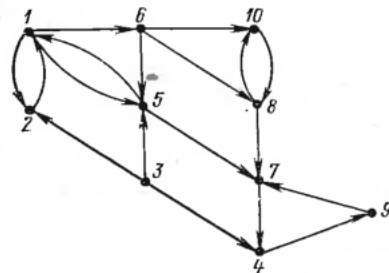


Рис. 2-7. Топологическая декомпозиция структуры

Изложенный метод декомпозиции основан только на анализе топологических свойств системы; функциональная сторона анализа в данном случае не рассматривается.

2.5. СТРУКТУРНО-ТОПОЛОГИЧЕСКИЕ ХАРАКТЕРИСТИКИ СИСТЕМ И ИХ ПРИМЕНЕНИЕ

При проведении структурного анализа систем очень часто необходимо располагать методикой, позволяющей определять некоторые структурные характеристики систем и давать им количественную оценку. Целесообразность определения таких характеристик состоит в том, что уже на ранней стадии проектирования появляется необходимость оценивать качество структуры системы и ее элементов с позиций общего системного подхода. Рассмотрим некоторые из них.

1. *Связность структуры.* Данная количественная характеристика позволяет выявить наличие обрывов в структуре, висящие вершины и др. Наиболее полно количественно связность элементов ориентированного графа определяется матрицей связности $C = \|C_{ij}\|$. Элементы матрицы C можно вычислить на основе матрицы $A_\Sigma = \sum_{k=1}^n A^k$. Элемент $C_{ij} = 1$, если $a_{ij}^\Sigma \geq 1$; $C_{ij} = 0$, если $a_{ij}^\Sigma = 0$. Для неориентированных графов связность всех элементов в структуре соответствует выполнению следующего условия:

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \geq n-1; i \neq j. \quad (2-10)$$

Правая часть неравенства определяет необходимое минимальное число связей в структуре неориентированного графа, содержащего n вершин.

2. *Структурная избыточность.* Структурный параметр, отражающий превышение общего числа связей над минимально необходимым, будем называть структурной избыточностью R .

В соответствии с (2-10) структурная избыточность R определяется следующим образом:

$$R = \frac{1}{2} \left[\sum_{i=1}^n \sum_{j=1}^n a_{ij} \right] \frac{1}{n-1} - 1.$$

Данная структурная характеристика используется для косвенной оценки экономичности и надежности исследуемых систем. Для систем с максимальной избыточностью, имеющих структуру типа «полный граф», $R > 0$; для систем с минимальной избыточностью $R = 0$; для систем несвязных $R < 0$.

Таким образом, система с большей избыточностью R потенциально более надежна, однако в ряде задач анализа структурной надежности, как показано в [23], ее целесообразно дополнить другим параметром, учитывающим неравномерность распределения связей, ϵ^2 .

Равномерное распределение связей в структуре неориентированного графа, имеющего m ребер и n вершин, характеризуется средней степенью вершины $\bar{\rho} = 2m/n$. Тогда, введя понятие отклонения $\rho_i - \bar{\rho}$, где ρ_i — действительная степень i -й вершины заданного графа, можно определить квадратическое отклонение заданного распределения степеней вершин от равномерного:

$$\begin{aligned} \epsilon^2 &= \sum_{i=1}^n (\rho_i - \bar{\rho})^2 = \\ &= \sum_{i=1}^n \rho_i^2 - 2\bar{\rho} \sum_{i=1}^n \rho_i + 4m^2/n = \\ &= \sum_{i=1}^n \rho_i^2 - 2 \cdot 2m/n \cdot 2m + 4m^2/n = \sum_{i=1}^n \rho_i^2 - 4m^2/n. \end{aligned}$$

Показатель ϵ^2 характеризует недоиспользование возможностей заданной структуры, имеющей m ребер и n вершин, в достижении максимальной связности. В [23] указанный показатель в относительном выражении использован для сравнения различных структур АСУ.

3. *Структурная компактность.* Для количественной оценки структурной компактности вводится параметр, отражающий близость элементов между собой. Близость двух элементов i и j между собой будем определять через минимальную длину пути для ориентированного графа (цепи — для неориентированного) d_{ij} . Тогда величина

$Q = \sum_{i=1}^n \sum_{j=1}^n d_{ij}, (i \neq j)$ отражает общую структурную близость элементов между собой в системе. Для количественной оценки структурной компактности очень часто исполь-

зуют относительный показатель

$$Q_{\text{отн}} = \frac{Q}{Q_{\text{мин}}} - 1,$$

где $Q_{\text{мин}} = n(n-1)$ — минимальное значение компактности для структуры системы типа «полный граф».

Структурную компактность можно характеризовать и другой характеристикой — диаметром структуры: $d = \max_{ij} d_{ij}$. Учитывая преобладающий информационный

характер связей в БСУ, можно сказать, что как величина $Q_{\text{отн}}$, так и d интегрально оценивают инерционность информационных процессов в системе, а при равных значениях ϵ^2 и R их возрастание отражает увеличение числа разделяющих связей в структуре, характеризуя тем самым снижение общей надежности.

4. *Степень централизации в структуре.* Для количественной оценки степени централизации в структуре используется понятие индекса центральности δ [23]:

$$\delta = (n-1) \frac{(2z_{\text{макс}} - n)}{(z_{\text{макс}}(n-2))},$$

где $z_{\text{макс}}$ — максимальное значение величины

$$z_i = \frac{Q}{2} \left(\sum_{j=1}^n d_{ij} \right)^{-1}, \quad i = 1, 2, \dots, n; \quad i \neq j.$$

Для структур систем, имеющих максимальную степень централизации (рис. 2-8, е), $\delta = 1$; для структур с равномерным распределением связей (рис. 2-8, б, д) $\delta = 0$.

5. *Ранг элемента* используется при представлении структуры системы в виде ориентированного графа. Данная характеристика позволяет распределить элементы системы в порядке их значимости. Значимость элемента здесь определяется только числом связей данного элемента с другими. Безусловно, ранг элемента в таком определении не является полной характеристикой важности элемента в системе, так как при этом не учитываются точностные, информационные и другие функциональные характеристики элемента. Однако, характеризуя значимость элемента рангом, можно высказать такое весьма правдоподобное допущение: чем выше ранг элемента, тем более сильно он связан с другими элементами системы и тем более тяжелыми будут последствия при изменении качества его функционирования. Строгое определение ранга элемента, при-

веденное в [37], сопряжено с определенными вычислительными трудностями, поэтому на этапе структурного анализа вполне можно ограничиться приближенным способом

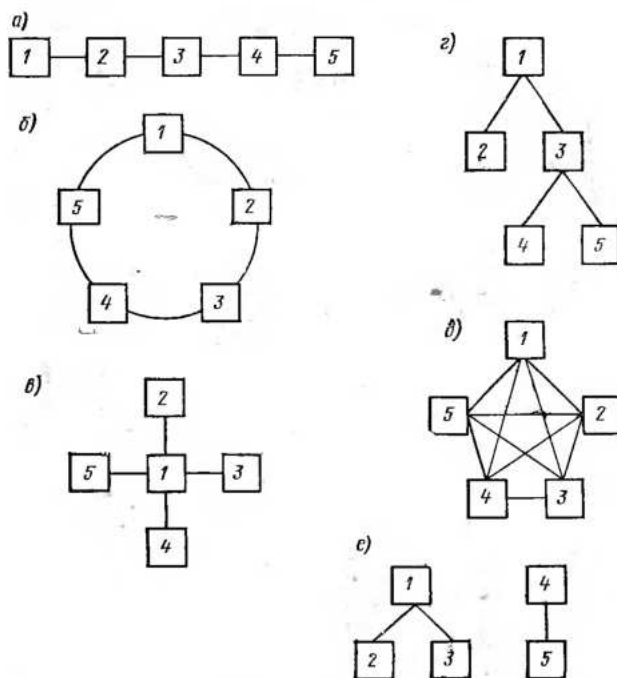


Рис. 2-8. Основные виды структур: а — последовательная; б — кольцевая; в — радиальная; г — древовидная; д — полный граф; е — несвязная

определения ранга элемента согласно следующей формуле:

$$r_i \approx \frac{\sum_{j=1}^n a_{ij}^{(k)}}{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^{(k)}},$$

где $a_{ij}^{(k)}$ — элементы матрицы A^k ; $k = 3 \div 4$.

Для практических задач такой способ дает близкие к истинным значения относительных рангов элементов и не требует больших вычислений. Значения рангов элементов являются весьма полезной информацией при распределе-

нии временных, стоимостных, технических ресурсов для достижения целей, поставленных при проектировании БСУ.

Введенные выше количественные характеристики могут быть использованы при сравнительной оценке топологических свойств структур систем. С точки зрения топологии внутренних связей выделяют следующие основные виды структур (рис. 2-8): а) последовательная; б) кольцевая; в) радиальная; г) древовидная; д) типа «полный граф»; е) несвязная. Рассмотрим применение количественных характеристик к анализу топологических свойств этих структур (число элементов равно пяти). Результаты анализа представлены в табл. 2-1.

Таблица 2-1

Вид	R	ϵ^2	Q	d	δ
а	0	1,2	1,0	4	0,7
б	0,25	0	0,5	2	0
в	0	7,2	0,6	2	1,0
г	0	3,2	0,7	3	0,7
д	1,50	0	0	1	0
е	-0,25	—	—	—	—

Из табл. 2-1 видно: 1) для несвязных структур $R < 0$; для структур без избыточности (последовательная, радиальная, древовидная) $R = 0$; для структур с избыточностью по связям (кольцевая, типа «полный граф») — $R > 0$; 2) структуры (последовательная, радиальная, древовидная) с $R = 0$ различаются по показателю ϵ^2 ; наибольшую неравномерность связей имеет радиальная структура; 3) наибольшую близость элементов (показатель Q) имеет структура типа «полный граф»; наименьшую — последовательная; радиальная и кольцевая структуры, неразличимые по показателю d , имеют различные значения Q ; 4) радиальная и древовидная структуры, имеющие одинаковые или близкие значения R , Q , d , значительно отличаются по показателям ϵ^2 и δ , что соответствует физическому смыслу, ибо отход от полной централизации в структуре ведет к большой равномерности распределения связей по элементам.

Рассмотренные выше структурные характеристики были получены только на основе информации о составе элементов и их связях. Дальнейшее развитие методологии построе-

ния структурных параметров для решения задач структурного анализа может быть основано на учете неструктурной информации за счет введения числовых функций на графах. Это позволяет наряду с составом элементов и направленностью их взаимодействия учитывать при решении задач структурного анализа другие интересные стороны взаимодействия (временные, надежностные, стоимостные и др.).

Рассмотрим далее третий уровень формализации задач структурного анализа, когда учитывается состав сигналов взаимосвязи элементов и их вид.

2-6. МОДЕЛЬ СТРУКТУРНОГО СОПРЯЖЕНИЯ — ЭЛЕМЕНТОВ В БСУ

Для построения математической модели сопряжения элементов в БСУ с учетом состава и характера сигналов взаимосвязи делается ряд допущений: 1) входной сигнал $x(t)$, поступающий к элементу в момент t , рассматривается как совокупность элементарных сигналов $x_1(t)$, $x_2(t)$, ..., $x_m(t)$, одновременно возникающих на входах элемента. Аналогично, выходной сигнал $y(t)$ — как совокупность элементарных сигналов $y_1(t)$, $y_2(t)$, ..., $y_n(t)$; 2) элементарные сигналы передаются в системе независимо друг от друга по элементарным каналам; 3) к входному контакту любого элемента подключается не более чем один элементарный канал; к выходному контакту может быть подключено любое конечное число элементарных каналов.

Данные допущения хорошо согласуются с характером взаимодействия реальных систем. Итак, пусть система S состоит из N элементов. Для i -го элемента C_i имеем $X^{(i)}$ — множество входных сигналов $x^{(i)}$, ($x^{(i)} \in X^{(i)}$), $Y^{(i)}$ — множество выходных сигналов $y^{(i)}$, ($y^{(i)} \in Y^{(i)}$).

Взаимодействие системы с внешней средой рассматривается как обмен сигналами между внешней средой и элементами системы. Каждый сигнал, выдаваемый во внешнюю среду, складывается из элементарных сигналов, выдаваемых одним или несколькими элементами системы; элементарные сигналы, составляющие сигнал, который поступает из внешней среды, принимаются одним или несколькими элементами.

В соответствии с этим внешнюю среду можно представить в виде фиктивного элемента системы C_0 , причем сигнал, выдаваемый системой во внешнюю среду, принима-

ется внешней средой как входной сигнал $x^{(0)}(t) = [x_1^{(0)}(t), x_2^{(0)}(t), \dots, x_{m_0}^{(0)}(t)]$, а сигнал, поступающий в систему из внешней среды, является выходным сигналом элемента, имитирующего внешнюю среду, и состоит из элементарных сигналов $y^{(0)}(t) = [y_1^{(0)}(t), y_2^{(0)}(t), \dots, y_{n_0}^{(0)}(t)]$.

Таким образом, для каждого элемента, включая и внешнюю среду, имеем множество входных контактов: $[X_j^{(i)}]_1^{m_i} = (X_1^{(i)}, X_2^{(i)}, \dots, X_{m_i}^{(i)})$ и множество выходных контактов: $[Y_j^{(i)}]_1^{n_i} = (Y_1^{(i)}, Y_2^{(i)}, \dots, Y_{n_i}^{(i)})$.

Введем некоторый оператор $Y_j^{(k)} = R(X_j^{(i)})$, где

$$X_j^{(i)} \in \bigcup_{i=0}^N [X_j^{(i)}]_1^{m_i}; Y_j^{(k)} \in \bigcup_{i=0}^N [Y_j^{(i)}]_1^{n_i},$$

противопоставляющий входному контакту $X_j^{(i)}$ выходной контакт $Y_j^{(k)}$, связанный с ним элементарным каналом, причем если в системе к контакту $X_j^{(i)}$ не подключен никакой элементарный канал, то оператор считается не определенным на этом $X_j^{(i)}$.

Оператор R называют оператором сопряжения. Обычно он задается в виде таблицы, в которой на пересечении строк с номерами элементов системы (i) и столбцов с номерами контактов (j) располагаются пары чисел (k, l), указывающие номер элемента (k) и номер контакта (l), с которым соединен контакт $X_j^{(i)}$. Для пояснения

Таблица 2-2

i	j		
	1	2	3
0	(3,1)	(4,1)	(4,2)
1	(0,1)	(0,2)	(3,2)
2	(0,3)	(0,4)	—
3	(1,1)	(1,2)	—
4	(1,3)	(2,1)	—

если контакты $(X_j^{(i)})$ и $(Y_j^{(k)})$ соединены элементарным каналом, то на пересечении соответствующей строки и столбца помещают единицу, в противном случае — нуль.

Хотя таблица такого рода громоздка и недостаточно наглядна, она обладает тем преимуществом, что представ-

данного принципа рассмотрим пример.

Пусть дана система (рис. 2-9). Таблица, задающая оператор сопряжения R , выглядит следующим образом:

В некоторых случаях оператор сопряжения R задают другим образом: столбцы и строки нумеруют двойными номерами (i, j) и (k, l) соответственно;

ляет собой матрицу смежности ориентированного графа, вершинами которого являются контакты, а дугами — элементарные каналы. Это позволяет применить рассмотренные методы теории графов для изучения структуры связей между входными и выходными контактами элементов системы и внешней средой.

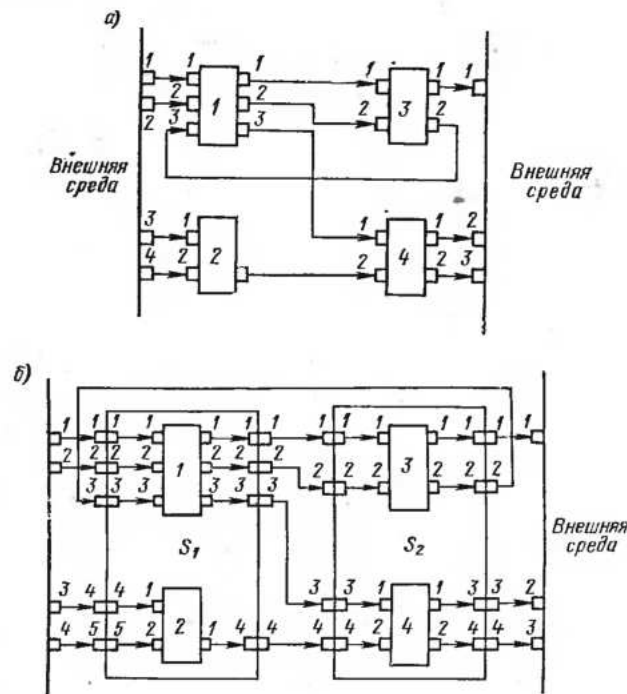


Рис. 2-9. Представление структуры системы: а — с одним уровнем; б — с двумя уровнями

Задание оператора R вышеуказанным образом определяет одноуровневую схему сопряжения. Приведенный способ формализации позволяет также выделять подсистемы и строить схемы сопряжения с большим числом уровней. В этом случае необходимо учитывать тот факт, что любая подсистема S_{ij} , с одной стороны, сама является системой, содержащей некоторое число элементов, а с другой — ее можно рассматривать как элемент анализируемой системы S .

Подсистема S_μ как самостоятельная система должна иметь контакты $X_j^{(0)\mu}$ и $Y_l^{(0)\mu}$, характеризующие источники и потребители внешней среды, а как элемент системы S подсистема S_μ должна содержать входные X_j^μ и выходные Y_l^μ контакты для связи с другими подсистемами. Соответствующие контакты $X_j^{(1)\mu}$ и $Y_l^{(0)\mu}$ при $j = l$, а также $X_j^{(0)\mu}$ и $Y_l^{(1)\mu}$ при $j = l$ объединяются в «двойные» контакты на границах подсистем S_μ . Пусть, например, система S состоит из двух подсистем S_1 и S_2 , причем в подсистему S_1 входят элементы C_1 и C_2 , а в подсистему S_2 — элементы C_3 и C_4 . В соответствии со сказанным выше подсистема S_1 как самостоятельная система имеет контакты $Y_1^{(0)1}, Y_2^{(0)1}, Y_3^{(0)1}, Y_4^{(0)1}, Y_5^{(0)1}; X_1^{(0)1}, X_2^{(0)1}, X_3^{(0)1}, X_4^{(0)1}$, а как элемент системы S имеет контакты $X_1^{(1)}, X_2^{(1)}, X_3^{(1)}, X_4^{(1)}, X_5^{(1)}; Y_1^{(1)}, Y_2^{(1)}, Y_3^{(1)}, Y_4^{(1)}$.

Аналогичным образом соответствующие множества могут быть построены и для подсистемы S_2 . Введем оператор $Y_l^{(k)} = R_\mu^{(1)}(X_j^{(0)})$, $\mu = 1; 2$ с областью определения

Таблица 2-3

i	j			
	1	2	3	4
0	(1,1)	(1,2)	(1,3)	(2,1)
1	(0,1)	(0,2)	(0,3)	—
2	(0,4)	(0,5)	—	—

Таблица 2-4

i	j			
	1	2	3	4
0	(3,1)	(3,2)	(4,1)	(4,2)
3	(0,1)	(0,2)	—	—
4	(0,3)	(0,4)	—	—

в подсистеме S_μ существует). Оператор $R_\mu^{(1)}$ называется внутренним оператором сопряжения подсистемы S_μ , ($\mu = 1; 2$). Способы задания оператора $R_\mu^{(1)}$ такие же, как и у обычного оператора сопряжения. Операторы сопряжения $R_1^{(1)}, R_2^{(1)}$ для подсистем S_1 и S_2 представлены соответственно в табл. 2-3, 2-4.

Операторы сопряжения $R_1^{(1)}$ и $R_2^{(1)}$ являются операторами сопряжения первого уровня. Рассмотрим теперь подсистему S_μ как элемент системы S . С этой точки она характеризуется множествами входных и выходных контактов: $[\hat{X}_j^\mu]_1^{m_\mu}$ и $[\hat{Y}_l^\mu]_1^{n_\mu}$. Элемент C_0 , представляющий внешнюю среду системы S , будем интерпретировать как подсистему S_0 с входными контактами $[X_j^{(0)}]_1^{m_0}$ и выходными контактами $[Y_l^{(0)}]_1^{n_0}$.

Введем оператор

$$\hat{Y}_l^{(k)} = R^{(2)}(\hat{X}_j^{(1)}),$$

реализующий отображение множества $\bigcup_{\mu=0}^M [\hat{X}_j^{(1)}]_1^{m_\mu}$ всех входных контактов подсистем $S_0, S_1, \dots, S_\mu$ системы S в множество $\bigcup_{\mu=0}^M [\hat{Y}_l^{(1)}]_1^{n_\mu}$ всех выходных контактов $\hat{Y}_l^\mu$,

который данному контакту $\hat{X}_j^{(1)}$ ставит в соответствие контакт $\hat{Y}_l^{(k)}$, соединенный с $\hat{X}_j^\mu$ элементарным каналом, если такое соединение с системе S существует. Оператор $R^{(2)}$ будем называть оператором сопряжения подсистем в системе S или оператором сопряжения второго уровня.

Оператор сопряжения $R^{(2)}$ подсистем S_1, S_2 и внешней среды S_0 представлен в табл. 2-5.

Таблица 2-5

i	j				
	1	2	3	4	5
0	(2,1)	(2,3)	(2,4)	—	—
1	(0,1)	(0,2)	(2,2)	(0,3)	(0,4)
2	(1,1)	(1,2)	(1,3)	(1,4)	—

Нулевая строка соответствует внешней среде, первая строка — подсистеме S_1 , вторая строка — подсистеме S_2 .

Совокупность внутренних одноуровневых схем сопряжения всех подсистем S_μ ($\mu = 1; 2$) вместе со схемой сопряжения второго уровня называется двухуровневой схемой сопряжения системы S .

Развитие подобного подхода легко обобщить на создание многоуровневых структурных моделей БСУ.

При анализе взаимодействия элементов в системе и с внешней средой из множества входных сигналов X для

любого элемента выделяется отдельно множество управляющих сигналов, воздействующих на его структуру и алгоритм функционирования. Поэтому все элементы, составляющие структуру системы, могут быть классифицированы следующим образом: 1) элементы, принимающие входные или управляющие сигналы от внешней среды (они называются входными полюсами); 2) элементы, выдающие выходные сигналы полностью или частично во внешнюю среду (они называются выходными полюсами); 3) элементы, принимающие входные или управляющие сигналы только от других элементов, входящих в систему (они называются внутренними).

По аналогии с понятиями пути и контура в теории ориентированных графов при данном подходе к описанию структуры вводятся следующие понятия. Канал $C_j C_k$ — такая последовательность элементов $C_j, C_{j+1}, \dots, C_k$, когда каждый предыдущий элемент является источником сигналов для последующего; если последовательность элементов начинается входным полюсом и заканчивается выходным полюсом, то канал называется сквозным; если элементы в последовательности $C_j, C_{j+1}, \dots, C_{k-1}, C_k$ обмениваются входными либо управляющими сигналами, причем C_{k-1} обязательно управляет элементом C_k , то канал $C_j C_k$ называется каналом управления; в противном случае канал называется каналом следования.

Контур — канал ($C_j C_k$) в случае, когда элемент C_k является источником сигналов для элемента C_j . Если выходные сигналы любого элемента в контуре воспринимаются только как входные, но не управляющие для следующего элемента, то контур называется контуром следования. Если выходные сигналы хотя бы одного из элементов, входящих в контур, воспринимаются следующим элементом в качестве управляющих, то контур называется контуром управления.

Введенные понятия канала и контура управления являются основой для выделения в структуре системы местных и общих каналов и контуров управления и определения составляющих элементов, охваченных ими.

В результате решения вышеуказанных задач структурного анализа на трех уровнях рассмотрения удастся: 1) оценить отношение между элементами и подсистемами БСУ, причем представить их в такой форме, что задачи анализа могут быть выполнены при помощи ЭВМ; 2) проследить

потоки информации, циркулирующие в системе; 3) выделить элементы и структурные конфигурации, существенные для функционирования системы и управления ею; 4) оценить качество структурной схемы, используя ее структурно-топологические и другие характеристики, что является необходимым этапом при решении задач макропроектирования БСУ. Далее рассмотрим основные подходы к построению моделей функционирования БСУ.

ГЛАВА ТРЕТЬЯ

АГРЕГАТИВНЫЕ МОДЕЛИ ФУНКЦИОНИРОВАНИЯ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

3-1. ОБЩАЯ МОДЕЛЬ ФУНКЦИОНИРОВАНИЯ

Формализация общей схемы процесса функционирования систем, в том числе и БСУ, основана на следующих положениях: 1) любая система функционирует во времени, взаимодействуя с внешней средой, и в каждый момент времени может находиться в одном из возможных состояний; 2) на вход системы могут поступать входные сигналы; 3) система способна выдавать выходные сигналы; 4) состояние системы в данный момент времени определяется предыдущими состояниями и входными сигналами, поступившими в данный момент времени и ранее; выходной сигнал в данный момент времени определяется состояниями системы и входными сигналами, относящимися к данному и предшествующим состояниям. Дадим формальную интерпретацию каждому из упомянутых положений.

1. Множество моментов времени t , в которые рассматривается функционирование системы, обозначим T . Множество T в общем случае является подмножеством множества действительных чисел и может быть непрерывным, дискретным (конечным или счетным) или дискретно-непрерывным. К системам, функционирующим в дискретном времени, обычно относят конечные автоматы, вычислительные устройства, ЭВМ. К системам, функционирующим в непрерывном времени, относятся многие электрические системы, системы автоматического управления непрерывными объектами, механические системы.

В ряде случаев функционирование систем носит дискретно-непрерывный характер, что характерно для иерархических систем автоматизированного управления технологическими процессами; на нижних уровнях управление может рассматриваться в непрерывном времени, а на более высоких — в дискретном времени. Функционирование системы во времени рассматривается как процесс перехода системы из состояния в состояние. Следовательно, прежде чем рассматривать функционирование системы, необходимо определить множество состояний Z . Возможны различные способы определения состояний. Рассмотрим некоторые из них.

А. Состояние системы определяется как совокупность состояний элементов системы. Если элементы системы могут находиться в двух состояниях, то система, составленная из n элементов, может находиться в одном из 2^n состояний. Так определяется состояние дискретных автоматов в вычислительной технике, сложных систем при анализе надежностных характеристик, кодирующих и декодирующих преобразователей в технике передачи данных и др. Таким образом,

$$z = (z_1, z_2, \dots, z_i, \dots, z_n); z \in Z,$$

где множество состояний системы z дискретно.

Б. Состояние системы характеризуется некоторым целым неотрицательным числом z ($z = 0, 1, 2, 3, \dots$). Так определяется состояние при анализе сложных информационных систем с одной фазой обслуживания. В этом случае z — число задач, запросов, находящихся в системе (на обслуживании либо в очереди).

В. Случае, обобщающим второй, является вариант, когда состояние системы описывается набором целых неотрицательных чисел:

$$z = (z_1, z_2, \dots, z_i, \dots, z_r); z_i = 0, 1, 2, \dots,$$

где z_i — число требований в i -й фазе; r — число фаз.

Подобный подход в определении состояния используется при описании динамики многофазных, многоэтапных систем типа телеавтоматических систем массового обслуживания.

Г. Состояние системы определяется набором действительных чисел. Например, положение самолета можно в данный момент времени описать вектором фазовых координат (z_1, z_2, z_3) , где z_1 — наклонная дальность, z_2 — азимут, z_3 — угол места.

В общем случае, охватывающем все рассмотренные варианты, будем считать, что состояние рассматриваемой БСУ z описывается некоторым набором характеристик z_i , $z_i \in Z_i$ ($i = 1, 2, \dots, k$), где Z_i — заданные множества, а множество возможных состояний БСУ определяется как прямое произведение множеств Z_i :

$$Z = Z_1 \times Z_2 \times \dots \times Z_i \times \dots \times Z_k.$$

Множество Z обычно в литературе называют пространством состояний системы. Аналогом пространства состояний в реальной физической интерпретации является трехмерное евклидово пространство. В общем случае состояние $z(t)$ в момент времени t есть точка пространства z с координатами $z_1, z_2, \dots, z_k$.

2. На вход любой сложной системы могут поступать входные сигналы $x \in X_0$, где X_0 — множество входных сигналов. Входной сигнал, поступающий в систему в момент $t \in T$, обозначим $x(t)$. В общем случае будем предполагать, что входной сигнал описывается набором характеристик $x_i \in X_i$ ($i = 1, 2, \dots, m$), где X_i — заданные дискретные или непрерывные множества.

Прямое произведение вида

$$X = X_1 \times X_2 \times \dots \times X_i \times \dots \times X_m$$

будем называть пространством входных сигналов, а входной сигнал x представляет собой точку пространства X , описываемую координатами $x_1, x_2, \dots, x_i, \dots, x_m$.

Множество входных сигналов $X_0 \subset X$; при этом следует считать, что множеству X_0 и принадлежит сигнал x_0 , означающий отсутствие сигнала в момент времени t . Отображение $x = L(t)$, противопоставляющее каждому $t \in T$ некоторый сигнал $x \in X_0$, будем называть входным процессом $L(t)$.

3. Система способна выдавать выходные сигналы $y \in Y_0$, где Y_0 — множество выходных сигналов. Выходной сигнал, выдаваемый системой в момент $t \in T$, обозначим $y(t)$. Если выходной сигнал y описывается набором характеристик $y_i, y_2, \dots, y_i, \dots, y_n$, таких, что $y_i \in Y_i$ ($i = 1, 2, \dots, n$), а Y_i — заданные множества, то прямое произведение

$$Y = Y_1 \times Y_2 \times \dots \times Y_i \times \dots \times Y_n$$

будем называть пространством выходных сигналов ($Y_0 \subset Y$). По аналогии с входным процессом определяется понятие выходного процесса $y = M(t)$.

Для определения поведения БСУ через введенные понятия множества состояний, множества входных, выходных сигналов необходимо ввести соотношения, позволяющие определить состояние системы и ее выходной сигнал в заданный момент времени t , если известны состояние системы и входные сигналы в момент времени, предшествующий моменту t .

В дальнейшем будем рассматривать динамику только для класса БСУ, называемых системами без последствия. Данный класс систем имеет существенное значение и в теории, и в практических приложениях, поэтому все методы описания и анализа динамики, рассматриваемые в последующих главах, будут ориентированы в основном на такие системы.

Система без последствия — система, будущее поведение которой определяется ее настоящим состоянием и не зависит от прошлого; иначе говоря, состояние $z(t)$ системы без последствия в момент времени $t > t_0$ определяется ее состоянием $z(t_0)$ и участком входного процесса $x = L(t)$ за интервал (t_0, t) , но не зависит от предыстории поведения системы, от того, каким образом система пришла в состояние $z(t_0)$. Различают два класса БСУ без последствия: детерминистические системы без последствия; стохастические системы без последствия.

Динамика для детерминистических систем будет считаться заданной, если указаны оператор переходов H , оператор выходов G .

1. Оператор переходов H определяет динамику переходов системы из состояния в состояние:

$$z(t) = H\{t_0, t, z(t_0), (t, x_L)_{t_0}^t\}, \quad (3-1)$$

где $z(t_0)$ — начальное состояние, $z(t_0) \in Z$; $t_0 \in T$, $t \in T$; $(t, x_L)_{t_0}^t \in \{(t, x_L)_{t_0}^t\}$; $\{(t, x_L)_{t_0}^t\}$ — множество всевозможных участков входного процесса $x = L(t)$, соответствующих интервалу (t_0, t) . При фиксированных t_0 , $z(t_0)$ и $(t, x_L)_{t_0}^t$ оператор H реализует отображение $z = H(t)$ или $z = z(t)$ множества T во множество Z , которое называется движением системы. Множество всевозможных движений системы обозначим $\{z(t)\}$.

Совокупность упорядоченных пар (t, z) для всех $t \in T$, где z определяется заданным движением $z = z(t)$, назы-

вается фазовой траекторией системы. Совокупность точек пространства Z , соответствующих в силу отображения $z = z(t)$ всем $t \in T$, называется траекторией системы в пространстве состояний.

Траектория системы является проекцией фазовой траектории на пространство Z . Оператор переходов H должен удовлетворять следующим двум основным условиям:

$$a) H\{t_0, t_0, z(t_0), (t, x_L)_{t_0}^{t_0}\} = z(t_0);$$

$$b) H\{t_0, t_2, z(t_0), (t, x_L)_{t_0}^{t_2}\} = \\ = H\{t_1, t_2, z(t_1), (t, x_L)_{t_1}^{t_2}\},$$

где $t_0 \in T$; $z(t_0) \in Z$; $z(t_1) \in Z$ и $(t_0 x_L)_{t_0}^{t_2}$ есть результат сочленения входных процессов $(t, x_L)_{t_0}^{t_1}$ и $(t, x_L)_{t_1}^{t_2}$.

Второе условие, называемое условием однозначности, позволяет разбивать входной процесс $x(t)$ на произвольные сочлененные между собой участки и анализировать поведение системы на каждом участке отдельно, что упрощает процедуру исследования систем.

2. Оператор выходов системы G определяет динамику выходных сигналов:

$$y(t) = G\{t, t_0, z(t_0), (t, x_L)_{t_0}^t\} = G[t, z(t)]. \quad (3-2)$$

Несмотря на внешнее сходство операторов, между ними имеется и существенное различие. Оператор H каждому $t > t_0$ из множества T ставит в соответствие определенный элемент $z \in Z$ (пространства состояний). Однако из физических соображений ясно, что существуют системы, которые выдают выходные сигналы не обязательно в каждый момент $t \in T$. Чтобы устранить это различие, по крайней мере с формальной точки зрения, можно предположить, что множеству Y_0 принадлежит и пустой сигнал y_ϕ , который физически интерпретируется как отсутствие выходного сигнала в момент времени t .

Для задания обобщенной динамики БСУ операторы H , G объединяют в виде оператора $F = H \times G$, обозначая его как оператор функционирования системы, а совокупность точек $[z(t), y(t)]$ пространства $Z \times Y$, соответствующих всем $t \in T$, определяют как траекторию функционирования.

Для задания динамики стохастических систем без последствия обычно вводят понятие случайного оператора, так как функционирование подобных систем

происходит при наличии случайных факторов. Пусть Ω — пространство элементарных случайных событий с вероятностной мерой $P(\omega)$, где ω — элементарное случайное событие. Тогда операторы переходов и выходов сложной стохастической системы без последствия обычно задаются в виде случайных операторов:

$$z(t) = H\{t, t_0, z(t_0, \omega_0), (t, x_L)_{t_0}^t, \omega_1\};$$

$$y(t) = G[t, z(t), \omega_2],$$

где $\omega_0, \omega_1, \omega_2$ — элементарные случайные события, которые независимо выбираются из Ω в соответствии с вероятностными мерами $P_{z_0}(\omega_0), P_z(\omega_1), P_y(\omega_2)$. При этом, если ω_1

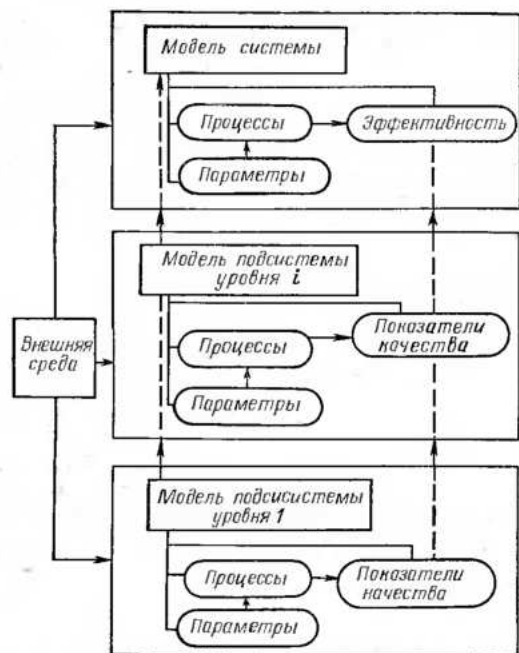


Рис. 3-1. Многоуровневое представление модели функционирования БСУ

и ω_2 фиксированы, то стохастическая система называется системой со случайными начальными состояниями; если ω_0 и ω_2 фиксированы, то стохастическая система называется системой со случайными переходами; если ω_0 и ω_1 фиксиро-

ваны, то стохастическая система называется системой со случайными выходами.

В силу наличия случайных факторов в системе динамику стохастической системы в пространстве состояний можно описывать поведением случайного процесса $z(t, \omega)$ с областью значений в множестве состояний системы Z .

В общем случае, как отмечено было в гл. 1, БСУ, в первую очередь, отличаются значительным числом элементов, их разнородным составом, различным характером взаимодействия. Поэтому построение общей модели функционирования в виде единых операторов H и G оказывается удобным с методической точки зрения, но затруднительным в решении практических задач.

Основной метод построения моделей функционирования БСУ состоит в функциональной декомпозиции системы, которая требует раскрытия структуры системы (состава элементов, функций элементов и их взаимодействия). Причем в большинстве случаев модель функционирования носит иерархический характер (рис. 3-1). Такой подход в построении моделей функционирования систем предложен и разработан Н. П. Бусленко в его работах и получил название агрегативного подхода [6].

3-2. АГРЕГАТИВНЫЕ МОДЕЛИ ЭЛЕМЕНТА И СИСТЕМЫ

Агрегат, предложенный в [6], является в некотором роде унифицированной моделью для описания функционирования разнородных элементов систем. Поэтому динамика БСУ раскрывается через динамику сопряженных между собой агрегативных моделей элементов. Способы задания сопряжения элементов были рассмотрены в § 2-6. Агрегат, как и система, характеризуется множествами моментов времени T , состояний в каждый момент времени Z , входных X_0 и выходных сигналов Y_0 . Рассмотрим особенности задания операторов H и G при агрегативном подходе. Для этого наряду с состоянием агрегата $z(t)$ введем состояние $z(t+0)$, в которое агрегат может перейти за малый интервал времени. Вид оператора H зависит от того, поступают или не поступают в течение рассматриваемого интервала времени входные сигналы. Пусть t_s — момент поступления в агрегат входного сигнала $x(t_s)$; тогда

$$z(t_s+0) = H_1[t_s, z(t_s), x(t_s)], t_s \in T.$$

Если полуинтервал $(t_s, t_{s+1}]$ не содержит ни одного момента поступления входных сигналов, за исключением t_{s+1} , то

$$z(t) = H_2[t, t_s, z(t_s + 0)], \quad t \in (t_s, t_{s+1}].$$

Во множестве состояний Z агрегата выделяют подмножество $Z^{(V)}$, такое, что если $z(t^*)$ достигает $Z^{(V)}$, то момент t^* является моментом выдачи выходного сигнала, определяемого оператором

$$y(t^*) = G[t^*, z(t^*)].$$

В некоторых случаях возможны изменения состояния в моменты выдачи выходных сигналов при выходе $z(t)$ на границу подмножества $Z^{(V)}$. Для учета этого факта вводится оператор H_3 :

$$z(t^* + 0) = H_3[t^*, z(t^*)].$$

Совокупность операторов H_1, H_2, H_3 задает ранее упомянутый оператор H , а операторы H и G полностью определяют модель функционирования агрегата. Процесс функционирования агрегата в основном состоит из скачков состояния в моменты поступления входных сигналов и выдачи выходных сигналов (операторы H_1, H_2) и изменений состояния между этими моментами (оператор H_3). Рассмотрим пример представления функционирования системы в виде агрегата.

Пример 3-1. Пусть в ЭВМ в случайные моменты времени t_j поступают задачи. Задача, поступившая в момент t_j , характеризуется случайным параметром a_j . Если задача застала вычислительную машину свободной, то задача принимается к решению; в противном случае задача направляется в очередь и находится там не более чем $v_j = \Phi(a_j, \beta)$, где β — параметр, характеризующий производительность ЭВМ. Если до момента времени $t_j + v_j$ задача j -й заявки не будет принята к обслуживанию, то она покидает систему. В момент окончания решения задачи ЭВМ приступает к решению следующей задачи в порядке очередности. Длительность решения задачи определяется как функция $\eta_j = \Psi(a_j, \beta)$. Состояние системы опишем следующими координатами: $z_1(t)$ — время, оставшееся до окончания решения задачи, находящейся на обслуживании; $z_2(t)$ — число задач в системе (в очереди и на обслуживании).

Если $z_2(t) = 0$ (в системе нет задач), то $z_1(t) = 0$ для всех $t \in T$ до момента появления новой задачи. Если $z_2(t) > 1$, то задачи имеются как на обслуживании, так и в очереди. В этом случае требуются дополнительные координаты состояния:

$$z_{1+2k}(t) = a'_k, \quad k = 1, 2, \dots, z_2(t) - 1,$$

где a'_k — параметр k -й задачи в очереди; $z_{2+2k}(t)$ — оставшееся время ожидания в очереди для k -й задачи.

Входные сигналы соответствуют задачам, поступающим в агрегат в моменты t_j , причем $x_j = x(t_j) = a_j$. Построим операторы H_1, H_2, H_3 для данного агрегата. Оператор H_1 , учитывающий изменение состояния системы в момент поступления входного сигнала при $z_2(t_j) > 0$, раскрывается следующим образом:

$$\begin{aligned} z_1(t_j + 0) &= z_1(t_j); \\ z_2(t_j + 0) &= z_2(t_j) + 1; \\ \left. \begin{aligned} z_{1+2k}(t_j + 0) &= z_{1+2k}(t_j); \\ z_{2+2k}(t_j + 0) &= z_{2+2k}(t_j); \end{aligned} \right\} k \leq z_2(t_j) - 1; \\ z_{1+2z_2(t_j)}(t_j + 0) &= a_j; \\ z_{2+2z_2(t_j)}(t_j + 0) &= \Phi(a_j, \beta), \end{aligned}$$

а при $z_2(t_j) = 0$ выглядит так:

$$\begin{aligned} z_1(t_j + 0) &= \Psi(a_j, \beta); \\ z_2(t_j + 0) &= 1, \end{aligned}$$

причем $z_{1+2k}(t_j + 0), z_{2+2k}(t_j + 0)$ — не определяются.

Рассмотрим изменение координат состояния в интервале между моментами поступления входных сигналов. После момента t_j координаты $z_1(t)$ и $z_{2+2k}(t)$ убывают с единичной скоростью, а $z_2(t)$ остается постоянной. Координата $z_1(t)$, убывая с единичной скоростью, обращается в нуль в момент t_i окончания обслуживания очередной задачи. В этот момент задача покидает систему и принимается к обслуживанию следующая $(i+1)$ -я задача из очереди, если $z_2(t_i) > 1$. Поэтому $z_1(t)$ от нуля скачком возрастает до $\eta_{i+1} = \Psi(a_{i+1}, \beta)$ и далее убывает с единичной скоростью. В этот же момент координата $z_2(t)$ уменьшается на единицу. Если же заявок в очереди нет, $z_1(t)$ остается равным нулю до поступления новой задачи и принятия ее к решению. Координаты $z_{2+2k}(t)$, убывая с единичной скоростью, обращались бы в нуль в моменты $t_j + v_j$ для соответствующих задач, если бы задачи не принимались к обслуживанию. Для тех задач, которые принимаются к обслуживанию, в моменты времени $t > t_i$ соответствующие координаты z_{2+2k} не определяются. В моменты времени $t_j + v_j$ (для задач, не принятых к обслуживанию) координата $z_2(t)$ уменьшается на единицу (задача покидает систему необслуженной). Прежде чем записывать соотношения для операторов H_2, H_3 , определим множество $Z^{(V)}$. В данном случае удобно выбрать множество $Z^{(V)}$ в виде объединения двух подмножеств $Z_1^{(V)}$ и $Z_2^{(V)}$. Подмножество $Z_1^{(V)}$ определяется соотношением $z_1(t) = 0$, а подмножество $Z_2^{(V)}$ — соотношением $z_{2+2k}(t) = 0$ хотя бы для одного значения k . Моменты t_1^* , для которых $z_1(t_1^*) = 0$, и моменты t_2^* , для которых $z_{2+2k}(t_2^*) = 0$, называются особыми и соответствуют моментам выдачи выходных сигналов.

Пусть $t = t_1^*$ (решение очередной задачи закончено). Здесь могут быть два случая. В первом (когда в системе имеются задачи, $z_2(t_1^*) > 1$) к обслуживанию принимается задача следующей заявки из очереди; время обслуживания $\tau = \Psi(a, \beta)$. Во втором случае (в системе нет задач, $z_2(t_1^*) - 1 = 0$) система ждет до момента поступления новой задачи и ее принимает на обслуживание. Оператор H_3 раскрывается

для этого случая следующим образом при $z_2(t_1^*) > 1$:

$$\begin{aligned} z_1(t_1^* + 0) &= \Psi(z_2, \beta); \\ z_2(t_1^* + 0) &= z_2(t_1^*) - 1; \\ z_{1+2k}(t_1^* + 0) &= z_{1+2(k+1)}(t_1^*); \\ z_{2+2k}(t_1^* + 0) &= z_{2+2(k+1)}(t_1^*); \end{aligned} \left. \vphantom{\begin{aligned} z_1(t_1^* + 0) &= \Psi(z_2, \beta); \\ z_2(t_1^* + 0) &= z_2(t_1^*) - 1; \end{aligned}} \right\} 1 \leq k \leq z_2(t_1^*) - 1,$$

а при $z_2(t_1^*) - 1 = 0$:

$$\begin{aligned} z_1(t_1^* + 0) &= z_1(t_1^*) = 0; \\ z_2(t_1^* + 0) &= 0; \\ z_{1+2k}(t_1^* + 0) \text{ и } z_{2+2k}(t_1^* + 0) \end{aligned}$$

не определяются.

Рассмотрим еще один особый момент времени t_2^* , не являющийся моментом поступления входного сигнала. В момент t_2^* , когда истекает время ожидания одной из задач, например l -й, число задач в системе уменьшается на единицу. Оператор H_3 для этого случая раскрывается следующим образом:

$$\begin{aligned} z_1(t_2^* + 0) &= z_1(t_2^*); \\ z_2(t_2^* + 0) &= z_2(t_2^*) - 1; \\ z_{1+2k}(t_2^* + 0) &= z_{1+2k}(t_2^*); \\ z_{2+2k}(t_2^* + 0) &= z_{2+2k}(t_2^*); \end{aligned} \left. \vphantom{\begin{aligned} z_1(t_2^* + 0) &= z_1(t_2^*); \\ z_2(t_2^* + 0) &= z_2(t_2^*) - 1; \end{aligned}} \right\} l < l; \\ z_{1+2k}(t_2^* + 0) &= z_{1+2(k+1)}(t_2^*); \\ z_{2+2k}(t_2^* + 0) &= z_{2+2(k+1)}(t_2^*); \end{aligned} \left. \vphantom{\begin{aligned} z_{1+2k}(t_2^* + 0) &= z_{1+2(k+1)}(t_2^*); \\ z_{2+2k}(t_2^* + 0) &= z_{2+2(k+1)}(t_2^*); \end{aligned}} \right\} l \leq k \leq z_2(t_2^*) - 1.$$

В полуинтервалах (t_n, t_{n+1}) между моментами t_n, t_{n+1} , к которым относятся моменты поступления входных сигналов и моменты типа t_1^* и t_2^* , состояния агрегата изменяются в соответствии с оператором H_2 , который раскрывается следующим образом:

$$\begin{aligned} z_1(t) &= z_1(t_n + 0) - (t - t_n); \\ z_2(t) &= z_2(t_n + 0); \\ z_{1+2k}(t) &= z_{1+2k}(t_n + 0); \\ z_{2+2k}(t) &= z_{2+2k}(t_n + 0) - (t - t_n). \end{aligned}$$

Выходные сигналы агрегата формируются в моменты t_1^* и t_2^* . Выходными сигналами агрегата будем считать совокупность характеристик задач, покидающих систему.

Пусть $y = (y_1, y_2)$, где y_1 — признак ($y_1 = 1$, если систему покидает обслуженная задача, и $y_1 = 0$, если — необслуженная), а y_2 — совокупность сведений о задаче (например, $y_2 = (a_j, \beta, t^*)$, означает, что задача поступила в систему с параметром a_j , обслуживаясь при значении параметра системы β , и «покинула» систему в момент t^*). В качестве y_2 в зависимости от конкретной задачи могут фигурировать и другие величины или функции их. Таким образом, действия оператора G сводятся к выбору признака y_1 и формированию сведений о заявке y_2 .

Следовательно, выражения $y(t_1^*) = (1, a_j, \beta, t_1^*)$, $y(t_2^*) = (0, a_j, \beta, t_2^*)$ задают оператор G для рассматриваемого агрегата. На этом будем считать законченным построение агрегата, описывающего функционирование данной системы.

Как показано в [6], предложенное агрегативное описание функционирования элемента оказывается достаточно универсальным. Различные математические модели функционирования элементов, формализуемые на языке теории дискретных автоматов, в виде систем дифференциальных уравнений, в виде моделей массового обслуживания на основе алгоритмического описания могут быть сведены к агрегативному представлению.

Таким образом, для создания агрегативной модели функционирования БСУ в целом необходимо: 1) разработать агрегативные модели элементов; 2) построить модель сопряжения агрегатов в соответствии с рассмотренными в § 2-6 принципами сопряжения. При учете управляемости агрегатов из множества входных сигналов для каждого агрегата отдельно выделяют множество управляющих сигналов.

Модель функционирования БСУ с учетом управления во многом зависит от принятых принципов управления. Рассмотрим далее основные свойства и особенности функционирования БСУ с учетом выбранной структуры управления.

3-3. УПРАВЛЕНИЕ В БОЛЬШИХ СИСТЕМАХ

Несмотря на то, что агрегативные связи по управлению в БСУ могут быть достаточно разнообразными, можно выделить основные классы структур систем с позиций управления.

Децентрализованная структура (рис. 3-2, а). Построение системы с такой структурой возможно только в случае независимости объектов управления по материальным, энергетическим, информационным и иным ресурсам. Для выработки управляющего воздействия на каждый объект необходима информация о состоянии только этого объекта. Фактически такая система представляет собой совокупность нескольких независимых систем со своей информационной и алгоритмической базой.

Централизованная структура (рис. 3-2, б). Такая структура предполагает реализацию всех процессов управления объектами в едином центральном органе управления, который осуществляет обработку информации об управляемых

объектах и выдает им непосредственно управляющие сигналы. Такая структура обладает рядом достоинств: 1) позволяет достаточно просто реализовать процессы информационного взаимодействия; 2) создает принципиальную возможность глобально-оптимального управления системой в целом; 3) исключает необходимость в пересылках промежуточных результатов; 4) позволяет легко корректировать

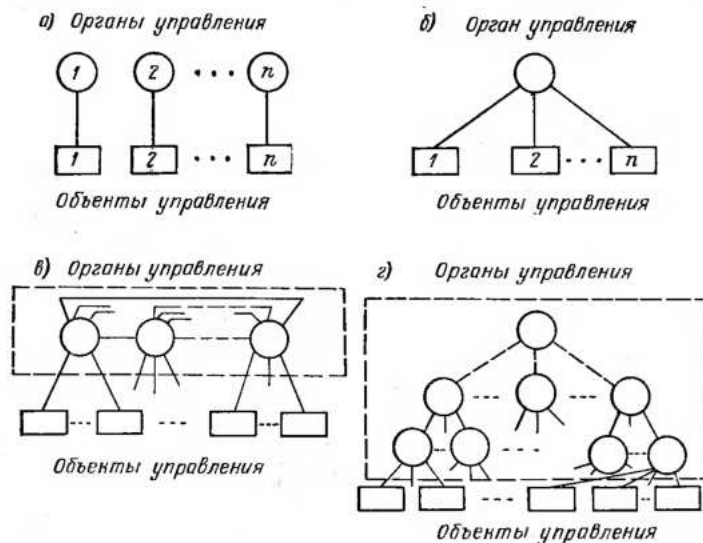


Рис. 3-2. Основные классы структур с позиций управления: а — децентрализованная; б — централизованная; в — централизованная рассредоточенная; г — иерархическая

оперативно-изменяемые данные; 5) дает возможность достигнуть максимальной эксплуатационной эффективности при минимальной избыточности технических средств.

С системотехнической точки зрения основными недостатками структуры систем с единым управлением являются: необходимость исключительно высокого объема запоминающих устройств, высокой производительности и надежности средств обработки данных для достижения приемлемого качества управления; высокая суммарная протяженность каналов связи при наличии территориально-рассредоточенных объектов управления.

Централизованная рассредоточенная структура (рис. 3-2, в). В указанной структуре сохраняется основная

особенность принципа централизованного управления — выработка управляющего воздействия на каждый объект на основе информации о состояниях всей совокупности объектов управления. Однако в отличие от полностью централизованной структуры процессов выработки управляющей информации она не имеет четкой локализации в едином управляющем органе. Алгоритм управления в этом случае состоит из совокупности взаимосвязанных алгоритмов управления объектами, которые реализуются совокупностью взаимно связанных органов управления.

В процессе функционирования каждый управляющий орган производит прием и обработку соответствующей информации, а также выдачу управляющих сигналов на подчиненные объекты. Для реализации функций управления каждый локальный орган по мере необходимости вступает в процесс информационного взаимодействия с другими органами управления. Достоинства такой структуры следующие: снижаются требования к объемам памяти, производительности, надежности каждого центра обработки и управления без ущерба для качества управления; суммарная протяженность и стоимость каналов связи в такой системе может быть снижена.

К недостаткам следует отнести: усложненность информационных процессов в системе вследствие необходимости обмена данными между центрами обработки и управления и корректировки хранимой информации; значительную избыточность технических средств, предназначенных для обработки информации, и, следовательно, повышенные расходы на их приобретение, размещение и эксплуатационное обслуживание; сложность синхронизации процессов обмена.

Иерархическая структура (рис. 3-2, г). С ростом масштабов объектов управления в БСУ чрезвычайно увеличивается необходимый объем перерабатываемой информации и сложность алгоритмов управления. В результате осуществлять управление централизованным образом (рис. 3-2, б) становится уже невозможно (хотя именно такое управление позволяет обеспечить наилучшую согласованность принимаемых решений), т. е. возникает необходимость разделения функций и органов управления. Такое разделение, позволяющее справиться с информационными трудностями для каждого местного органа управления, порождает необходимость согласования принимаемых этими органами решений, т. е. создания над ними нового управляющего органа. Таким образом, по мере роста сложности систем выстраивается

иерархическая пирамида управления. Поэтому главной причиной иерархического управления является несоответствие между сложностью управляемого объекта и способностью любого управляющего органа получать и перерабатывать информацию.

В БСУ, как и в любой системе управления, выделяют объект управления и органы управления. Многие объекты управления в БСУ имеют собственную иерархию. Особенно это относится к различным производственным системам. Иерархическая организация таких систем, т. е. образующих их производственных объектов и связей, возникла не из потребностей управления, а под влиянием объективных тенденций научно-технического прогресса — концентрации и специализации производства и централизации распределения продукции, которые способствуют повышению эффективности общественного производства. Поэтому концентрация, специализация производства и централизация распределения — это объективные системообразующие факторы, определяющие иерархию производственных объектов управления в больших автоматизированных системах. Одновременно с этим управляющая система имеет собственную иерархическую структуру, которая зачастую не совпадает с иерархией объекта управления. Управляемые процессы в сложном объекте управления требуют своевременного формирования правильных решений, которые, во-первых, приводили бы к поставленным целям, во-вторых, принимались бы своевременно и, в-третьих, были бы взаимно согласованы. Каждое такое решение требует постановки соответствующей задачи управления. Их совокупность образует иерархию задач управления, которая в ряде случаев значительно сложнее иерархии объекта управления. Решение этой иерархии задач управления является главной функцией органов управления.

Каждый отдельный орган может решать разное, но каждый раз — ограниченное число задач, т. е. имеет ограниченную пропускную способность. Поэтому иерархия задач порождает иерархию органов управления, которая может быть значительно сложнее иерархии задач управления и иерархии организационной структуры объектов управления (рис. 3-3).

Рассмотрим на примере интегрированной системы управления производством содержание и иерархию целей и задач управления (рис. 3-4). Рассматриваемая система включает в себя подсистемы 1-, 2- и 3-го уровней и построена по

иерархическому принципу. Подсистема 1-го уровня обслуживает производственный процесс. В подсистему входят в общем случае средства автоматизации местного значения, включающие в себя управляющие вычислительные машины (УВМ) и местные регуляторы (Р), которые, получая информацию от преобразователей (П), воздействуют на производственный процесс через соответствующие исполнительные механизмы (ИМ). Применение подсистемы 1-го уровня позволяет решить следующие задачи: стабилизацию управляющих параметров и поддержание их на заданном уровне; контроль за состоянием производственных объектов.

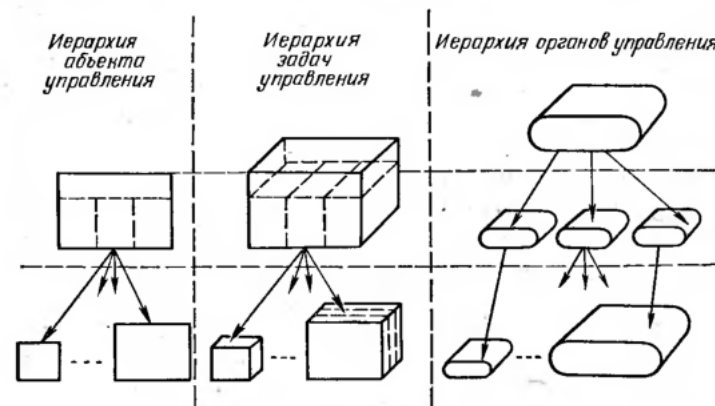


Рис. 3-3. Соотношение иерархий объекта управления, задач управления, органов управления

При этом возможно непосредственное воздействие подсистемы 2-го уровня на производственные объекты. Подсистема 2-го уровня в данном случае выступает как промежуточная и включает в себя подсистему сбора и передачи информации (ПСПИ2), машину УВМ2, подсистему выдачи данных (ПВД2). Машина УВМ2, обеспечивая воздействие на УВМ1 нижней ступени, позволяет решать целый ряд задач по управлению производственными участками: определение зоны оптимальных режимов работы соответствующего технологического оборудования; выдачу обобщенных показателей хода производства и передачу их в подсистему 3-го уровня с помощью ПВД2; выработку необходимых рекомендаций для управления производственными участками и передачу этих рекомендаций в подсистему 1-го уровня. Высшим органом — органом управления —

является подсистема 3-го уровня, которая координирует работу автоматических систем 1-го уровня и системы управления 2-го уровня. Подсистема 3-го уровня включает в себя ЭВМ, которая по своей производительности должна превосходить УВМ управляемых подсистем, ПСПИЗ и ПВДЗ,

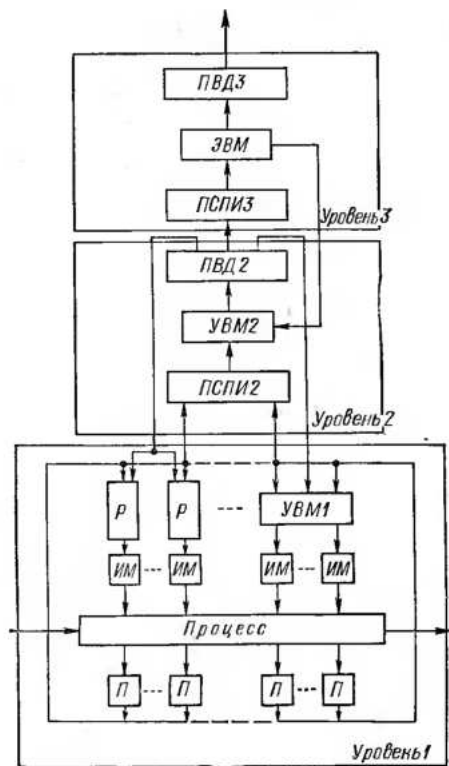


Рис. 3-4. Структура автоматизированной интегрированной системы управления производством

причем последняя далее выдает обобщенные данные о ходе производства непосредственно в дирекцию, в отделы управления либо в отраслевую автоматизированную систему управления. В основном подсистема 3-го уровня решает задачи, связанные с оптимизацией работы всего предприятия исходя из заданного критерия эффективности функционирования. Одновременно в подсистему 3-го уровня вводятся

экономическая и оперативно-статистическая информация, которая позволяет следить за выполнением плана производства и вовремя вмешиваться непосредственно в процесс производства.

Таким образом, в рассматриваемой БСУ выделяют три уровня задач управления: уровень 1 — стабилизация управляющих параметров и поддержание их на заданном уровне, контроль за состоянием производственных объектов; уровень 2 — управление производственными участками для обеспечения оптимальных режимов работы соответствующего технологического оборудования; уровень 3 — оптимизация работы всего предприятия, исходя из заданного критерия качества функционирования с выдачей соответствующих обобщенных данных о ходе производства в дирекцию, в отделы управления либо в отраслевую автоматизированную систему управления (ОАСУ).

Первая попытка сформулировать основные понятия, свойства и принципы управления в иерархических системах и создать фундамент для дальнейших исследований была сделана М. Месаровичем в его монографии «Теория иерархических многоуровневых систем». Из нее следует, что всем большим иерархическим системам управления присущи следующие основные особенности: последовательное вертикальное расположение подсистем, составляющих данную систему, или, точнее, вертикальная соподчиненность; приоритет действий или право вмешательства подсистем верхнего уровня в действие подсистем нижнего уровня; зависимость действий подсистем верхнего уровня от фактического исполнения подсистемами нижних уровней своих функций.

Организация многоуровневой структуры управления имеет следующие общие функциональные особенности.

1. *Элементы верхнего уровня управления* имеют дело с более крупными подсистемами или с более широкими аспектами поведения системы в целом и, следовательно, не могут реагировать на такие изменения в окружающей среде или самом объекте, с которыми имеют дело элементы нижних уровней.

2. *Чем выше уровень управления*, тем больше период принятия решения по формированию управления. Это связано с тем, что степень неопределенности информации о состоянии системы повышается с ростом номера уровня. Степень неопределенности зависит от состояния внешней среды, от стохастического характера функционирования системы, от индивидуальных характеристик людей, участвующих в уп-

равлении, от многозначности критериев оценки качества управления и др. В связи с этим за малый период управления не успевает накопиться необходимое количество информации, достаточной для формирования правильного управляющего воздействия. Кроме того, если длительность периода мала, то управляющее устройство может только реализовать упрощенные управляющие алгоритмы, соответствующие более грубым методам переработки информации и принятия решений.

3. Описание задач управления на верхних уровнях содержит больше неопределенностей и более трудно для количественной формализации, что затрудняет поиск наилучших управлений. Поэтому в тех случаях, когда задачи верхнего уровня не удается поставить так, чтобы они имели простое численное решение, на практике обычно прибегают «к вмешательству» верхних уровней в работу нижних только в критических ситуациях: структурные изменения вносятся тогда, когда характеристики ухудшаются до такой степени, что изменения становятся необходимыми.

В многоуровневой иерархии принятия решений по формированию управления в БСУ выделяют обычно три уровня: уровень выбора (нижний) — выбор способа действия в соответствии с исходной информацией о внешней среде и командными алгоритмическими предписаниями, поступающими с вышестоящего уровня; уровень обучения, адаптации (средний) — конкретизация множества неопределенностей для первого уровня за счет изучения внешней среды и определение вида алгоритма либо параметров алгоритма работы нижестоящего уровня; уровень самоорганизации (верхний) — выбор критериев, алгоритмов, их структуры, используемых на нижележащих уровнях таким образом, чтобы по возможности приблизиться к главной цели управления.

Наиболее распространенной в настоящее время является иерархическая структура с двумя уровнями управления, которую иногда называют централизованной структурой с автономным управлением (рис. 3-5). Рассмотрим характерные особенности иерархических структур БСУ на данном примере. Особенностью данной структуры является сочетание централизованного управления системой с локальным управлением отдельными объектами (либо группами). Таким образом, функции управления разделены между центральным органом, который реализует алгоритм глобального управления системой, и локальными органами,

каждый из которых решает задачи управления некоторым подмножеством объектов в соответствии с управляющими воздействиями центрального органа.

В рассматриваемой структуре сочетаются многие положительные функциональные свойства полностью централизованной и централизованной рассредоточенной структур: 1) максимальная автономность локальных центров в процессе управления с возможностью оптимального управления системой в целом; 2) централизованное хранение

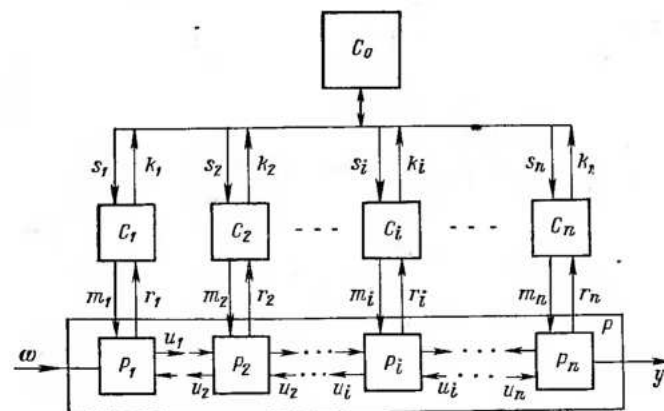


Рис. 3-5. Иерархическая структура с двумя уровнями управления

и обработка информации, относящейся ко всей системе в целом, с децентрализованным размещением и обработкой информации, необходимой для автономного управления отдельными объектами; 3) пониженные требования к пропускной способности и производительности локальных органов управления с высокой общей пропускной способностью (производительности системы); 4) система обладает повышенной надежностью, так как при выходе из строя центрального органа локальные центры управления могут продолжать функционировать, например, в соответствии с последним централизованным управляющим воздействием.

Рассмотренные выше достоинства обеспечили исключительно большую распространенность таких иерархических структур в больших автоматизированных системах управления типов АСУ ТП, АСУП.

Для анализа основных особенностей управления в иерархических системах дадим общесистемное описание функционирования двухуровневой системы. Система содержит: 1) вышестоящую управляющую систему C_0 ; 2) n нижележащих управляющих систем $C_1, C_2, \dots, C_n$; 3) управляемый процесс P . В системе выделены два вида вертикального взаимодействия между подсистемами. Один — это передача вниз командных сигналов; сигналы от управляющих систем $C_1, C_2, \dots, C_n$ к управляемому процессу P будем называть управляющими воздействиями, сигналы от вышестоящей C_0 к нижестоящим управляющим системам C_i будем называть координирующими сигналами или вмешательствами.

Так как системы нижнего уровня управления могут иметь различные локальные цели и, следовательно, находиться в конфликтных ситуациях, то задача координации для системы C_0 состоит в воздействии на нижестоящие системы таким образом, чтобы достигалась общая цель, заданная для всей системы в целом. Другой вид вертикального взаимодействия — это передача наверх информационных сигналов или сигналов обратной связи различным управляющим системам иерархии. Данный вид взаимодействия обеспечивает возможность того, что любая управляющая подсистема имеет сведения о ходе протекания самого процесса в объекте управления и о качестве управления.

Введем следующие обозначения: $m \in M$ — множество управляющих сигналов; $\omega \in \Omega$ — множество внешних возмущений; $r \in R$ — множество информационных сигналов 1-го уровня; $s \in S$ — множество координирующих сигналов; $y \in Y$ — множество выходных сигналов; $k \in K$ — множество информационных сигналов 2-го уровня.

Модель функционирования двухуровневой системы в данном случае будем представлять на языке теории множеств. Управляемый процесс задается отображением

$$P: M \times \Omega \rightarrow Y,$$

где $M = M_1 \times M_2 \times \dots \times M_i \times \dots \times M_n$ и M_i — множество управляющих сигналов для i -го управляющего органа [$m_i \in M_i$, $m = (m_1, m_2, \dots, m_n)$].

Модель функционирования i -й локальной системы управления ($i = 1, 2, \dots, n$) реализуется в виде отображения

$$C_i: S \times R_i \rightarrow M_i; R = R_1 \times R_2 \times \dots \times R_i \times \dots \times R_n.$$

Модель координатора выглядит как

$$C_0: K \rightarrow S.$$

Соответственно информационные обратные связи 1- и 2-го уровней реализуются отображениями:

$$f_i: M_i \times \Omega \times Y \rightarrow R_i; i = 1, 2, \dots, n;$$

$$f_0: S \times R \times M \rightarrow K.$$

В дальнейшем при анализе особенностей функционирования и принципов управления двухуровневой системы будем считать: 1) влияние внешней среды (возмущения) либо полностью отсутствует, либо полностью заранее известно, т. е. система в целом функционирует в условиях определенности; 2) общая задача управления и частные задачи управления, решаемые на нижележащем уровне, являются задачами оптимизации.

Глобальная задача оптимизации D_0 отражает глобальную цель управления двухуровневой системы и сводится к следующей постановке:

$$\text{найти } \min_{m \in M} g(m) = G[m, P(m)];$$

$$m = (m_1, m_2, \dots, m_n); M = M_1 \times M_2 \times \dots \times M_n.$$

Функция $g(m)$ является глобальной целевой функцией.

При рассмотрении локальных оптимизированных задач предполагается, что управляемый процесс P представлен в виде композиции подпроцессов P_i ($i = 1, 2, \dots, n$), взаимодействующих между собой, причем управление каждым подпроцессом P_i находится в ведении соответствующего управляющего элемента нижнего уровня C_i .

Взаимодействие подпроцесса P_i с другими реализуется через множество связей $u_i \in U_i$. Пусть D_i — локальная оптимизационная задача, решаемая i -м управляющим органом нижнего уровня, а локальная оптимизационная функция качества решения данной задачи выглядит следующим образом:

$$g_i(m_i, u_i) = G_i[m_i, P_i(u_i, m_i)],$$

где $M_i \times U_i$ — множество задания целевой функции.

Принципы координации являются центральным звеном при построении иерархических систем управления. Имеются два способа воздействия на локальные задачи оптимизации со стороны верхнего уровня: 1) через функцию качества G_i — координация путем изменения целей; 2) через изменение

параметров множеств связей U_i ($i = 1, 2, \dots, n$) в выделенном классе подпроцессов P_i — координация путем изменения ограничений.

Первый способ предполагает задание не одной, а множества локальных функций качества, в результате чего координирующий сигнал s_i направлен на выбор соответствующей функции качества функционирования из заданного числа i -й управляющей системы. В том случае, когда нет возможностей для координации путем изменения целей, используют координацию путем изменения ограничений на основе: 1) принципа развязывания взаимодействий u_i ; 2) принципа прогнозирования взаимодействий u_i .

Принцип развязывания взаимодействий предполагает, что каждый нижестоящий управляющий элемент получает право при решении собственной задачи управления рассматривать связующие входы u_i как дополнительные переменные, которые он выбирает из собственных локальных критериев. Очевидно, что подлежащие решению задачи управления нижнего уровня определяются в этом случае так, как если бы нижестоящие элементы и подпроцессы были полностью «развязанными» (т. е. автономными).

Принцип прогнозирования взаимодействий предполагает, что координирующие сигналы s_i содержат информацию о прогнозируемых значениях связей u_i , которые будут иметь место при подаче управляющих воздействий. Для иллюстрации рассматриваемых принципов координации рассмотрим примеры, считая, что задачами управления являются задачи оптимизации.

Пример 3-2. Задана двухуровневая система, имеющая на верхнем уровне один, а на нижестоящем уровне — два элемента. Пусть процесс P определяется системой двух уравнений

$$\left. \begin{aligned} y_1 &= m_1 + m_2; \\ y_2 &= -m_1 + m_2. \end{aligned} \right\}$$

Пусть далее функция качества управления определяется выражением

$$G(m, y) = m_1^2 + m_2^2 + (y_1 - 1)^2 + (y_2 - 1)^2.$$

Будем считать, что задачей управления является задача оптимизации, т. е. надо найти

$$\min_m G(m, P(m)).$$

Будем считать, что процесс P представлен в виде композиции двух взаимосвязанных процессов $P_1 = P_1(m_1, u_1)$ и $P_2 = P_2(m_2, u_2)$:

$$\left. \begin{aligned} y_1 &= 2m_1 + u_1; \\ y_2 &= 2m_2 - u_2, \end{aligned} \right\}$$

где $u_1 = y_2$, $u_2 = y_1$ определяют связи между процессами, причем управление первым процессом осуществляется первым управляющим элементом C_1 , а управление вторым процессом — вторым C_2 .

Пусть i -й управляющий элемент нижнего уровня вырабатывает управление m_i в соответствии с некоторой локальной функцией G_i :

$$G_1(m_1, u_1, y_1, s) = m_1^2 + (y_1 - 1)^2 + s_1 u_1^2 - s_2 y_1^2;$$

$$G_2(m_2, u_2, y_2, s) = m_2^2 + (y_2 - 1)^2 + s_2 u_2^2 - s_1 y_2^2.$$

Будем считать, что обе задачи управления, решаемые управляющими элементами нижнего уровня, представляют собой также задачи оптимизации, т. е. необходимо найти

$$\min_{u_i, m_i} G_i(m_i, P_i(m_i, u_i), s), \quad i = 1; 2$$

для любой заданной пары координирующих сигналов $s = (s_1, s_2)$. В соответствии с этим каждый управляющий элемент C_i ($s = 1; 2$) вырабатывает локальные управляющие сигналы $m_1(s)$, $m_2(s)$. Задача координатора C_0 сводится к отысканию такого значения s , чтобы соответствующие локальные управляющие воздействия $m_1(s)$, $m_2(s)$ были глобально-оптимальными управлениями, т. е. чтобы управляющее воздействие $m(s) = [m_1(s), m_2(s)]$ было решением глобальной задачи оптимизации. Заметим также, что при решении локальных задач оптимизации управляющие элементы нижестоящего уровня должны произвести выбор не только своих (локальных) управляющих воздействий, но и определенных значений связей u_1 и u_2 . Для каждого значения s $u_1^*(s)$ и $u_2^*(s)$ будут оптимальными значениями связей, т. е. их значениями, найденными в процессе оптимизации в предположении, что подпроцессы являются автономными и независимыми друг от друга. Однако фактически в силу существования взаимосвязи между подпроцессами такой свободы нет. Поэтому, если найденные независимые управляющие воздействия $m_1(s)$ и $m_2(s)$ действительно будут приложены к процессу P , то вследствие взаимосвязи подпроцессов фактические значения связей будут равны $u_1 = -m_1(s) + m_2(s)$, $u_2 = m_1(s) + m_2(s)$ и, вообще говоря, могут не совпадать со значениями, выбранными путем оптимизации.

В данном случае, если допустить, что $u_1 = u_1^*(s)$, $u_2 = u_2^*(s)$, т. е. выбираемые на нижележащем уровне значения связей подпроцессов P_1 , P_2 являются теми, что будут иметь место при приложении к процессу управляющих воздействий $m_1(s)$, $m_2(s)$, то управляющее воздействие $m(s) = [m_1(s), m_2(s)]$ является глобально-оптимальным. Если $s = (0; -1/2)$, то при оптимизации на низшем уровне имеем: $[m_1(s), u_1^*(s)] = (0; 2/3)$; $[m_2(s), u_2^*(s)] = (2/3; 2/3)$, а на высшем уровне $m(s) = (0; 2/3)$. Так как $u(s) = (2/3; 2/3)$, что совпадает со значениями связей, выбранными в процессе оптимизации, то найденное управление $m(s) = (0; 2/3)$ при координирующем сигнале $s = (0; -1/2)$ является глобально-оптимальным. Данный принцип координации называется принципом развязывания взаимодействия.

Пример 3-3. Задана двухуровневая система такой же структуры, как и в примере 3-2. Процесс P есть совокупность двух процессов P_1 , P_2 :

$$\left. \begin{aligned} y_1 &= m_1 + \frac{7}{3} y_2^3; \\ y_2 &= m_2 - y_1^3. \end{aligned} \right\}$$

Пусть M_i для $i = 1; 2$ есть множество вещественных чисел, таких, что $|m_i| \leq 1$. Глобальная функция качества $G(m, y)$ задана в следующем виде:

$$G(m, y) = m_1^2 + 4m_2^2 + y_1 + y_2.$$

Функции качества для локальных управляющих элементов C_1 и C_2 заданы

$$G_1(m_1, y_1) = m_1^2 - \frac{1}{3} m_1 + y_1;$$

$$G_2(m_2, y_2) = 4m_2^2 + \frac{1}{7} m_2 + y_2,$$

причем связь подпроцессов P_1 и P_2 аналогична той, что в примере 3-2, т. е. $u_1 = y_2$, $u_2 = y_1$.

Будем координировать задачи управления на нижележащем уровне, обязав координатор C_0 прогнозировать конкретные значения связей u_1 , u_2 . Глобальная и локальная задачи управления являются задачами минимизации. Координатор будет задавать прогнозируемые значения $s = (s_1, s_2)$ связей u_1 и u_2 . Затем управляющие элементы нижележащего уровня выберут локальные управляющие воздействия $m^*(s) = [m_1^*(s), m_2^*(s)]$, решая задачи оптимизации для прогнозируемых значений связей. Фактические связи будут иметь вид: $(u_1^*, u_2^*) = (y_1^*, y_2^*) = p[m^*(s), m_2^*(s)]$ — и, вообще говоря, могут не совпадать с прогнозируемыми значениями s_1, s_2 . Координация по принципу прогнозирования взаимодействия подразумевает: управляющее воздействие $m^*(s) = [m_1^*(s), m_2^*(s)]$ является глобально-оптимальным всякий раз, когда прогноз $s = (s_1, s_2)$ оправдывается, т. е. когда $u_1^* = s_1$, $u_2^* = s_2$. В данном случае принцип прогнозирования взаимодействия применим. Если выбрать $s_1 = -1/7$ и $s_2 = -1/3$, то для результирующих локальных управляющих воздействий имеем $m_1^*(s) = -1/3$, $m_2^*(s) = -1/7$, причем $m^*(s) = (-1/3, -1/7)$ является глобально-оптимальным управлением, так как фактические значения связей, получающиеся при применении $m^*(s)$: $u_1 = -1/7$, $u_2 = -1/3$.

Расхождение между прогнозируемыми и фактическими значениями связей могут быть использованы для того, чтобы судить о правильности выбранного координирующего воздействия. Кроме того, различия между прогнозируемыми и фактическими входами могут служить для оценки того, насколько далек от оптимума режим работы системы.

Содержание обоих примеров показывает, что координируемость в двухуровневой иерархической системе управления (оптимизации) означает существование оптимального координирующего сигнала, обеспечивающего экстремум глобальной функции качества управления при определенной декомпозиции общей задачи управления на ряд задач управления, решаемых на различных уровнях. Причем выбор принципов координации и способов формирования координирующих сигналов существенно зависит от принятой декомпозиции управляемого процесса P . Теория координации в иерархических системах составляет содержание новой интенсивно развивающейся области управления [34].

Таким образом, рассмотренный в данной главе агрегативный подход к построению описания функционирования БСУ хотя и является достаточно общим и универсальным, однако обладает определенной ограниченностью при аналитическом исследовании различных свойств БСУ. Примене-

ние аналитического аппарата всегда связано с определенными допущениями и упрощениями, что само по себе является отходом от универсальности и возвращением на тот уровень описания, который для данной системы и данных задач исследования позволяет построить более простую модель, адекватную поведению системы. Однако при исследовании поведения систем на ЭВМ путем моделирования агрегативный подход оказывается весьма перспективным направлением, так как позволяет разработать программные модели имитации функционирования для широкого класса систем. Так как функционирование БСУ является достаточно многогранным, то при решении задач макропроектирования приходится строить обычно не одну, а множество аналитических, численных, программных моделей для исследования различных функциональных свойств и характеристик (производительности, надежности, точности, качества управления и др.), объединенных в общую модель оценки показателей эффективности.

В последующих главах излагаются некоторые новые и уже хорошо зарекомендовавшие себя способы анализа различных функциональных свойств БСУ, применяемые на этапе макропроектирования систем (выбора структуры, алгоритмов управления, параметров взаимодействия систем).

ГЛАВА ЧЕТВЕРТАЯ

ПРИМЕНЕНИЕ МАРКОВСКИХ ПРОЦЕССОВ ДЛЯ АНАЛИЗА ПОВЕДЕНИЯ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

4-1. МАРКОВСКИЕ ПРОЦЕССЫ С КОНЕЧНЫМ ИЛИ СЧЕТНЫМ МНОЖЕСТВОМ СОСТОЯНИЙ И НЕПРЕРЫВНЫМ ВРЕМЕНЕМ

Класс марковских случайных процессов является простейшим и наиболее исследованным в теории случайных процессов. В настоящее время он весьма широко используется при анализе характеристик БСУ для описания их поведения при наличии случайных факторов. В дальнейшем из всего многообразия случайных марковских процес-

сов будут рассматриваться только процессы с конечным или счетным множеством состояний.

Пусть имеется дискретный случайный процесс, протекающий в системе с возможными состояниями: $z_0, z_1, z_2, \dots, z_i, \dots, z_j, \dots$. Обозначим условную вероятность того, что в момент $t = t_0 + \tau$ система будет в состоянии z_j , если в момент t_0 она была в состоянии z_i , через $P_{ij}(t_0, \tau)$. Дискретный случайный процесс функционирования системы называется марковским, если вероятность $P_{ij}(t_0, \tau)$ зависит только от указанных в обозначении параметров i, j, t_0, τ , т. е. от того, в каком состоянии была система в момент t_0 и в какое состояние она может перейти через время τ . Различают два основных вида марковских случайных процессов: с дискретным и с непрерывным временем перехода. Марковским случайным процессом с дискретным временем называется процесс, у которого переходы из одного состояния в другое возможны в строго определенные моменты времени.

Марковским случайным процессом с непрерывным временем называется процесс, у которого переход из одного состояния в другое возможен в любой момент времени. Марковские случайные процессы с непрерывным временем очень широко используются в практике анализа характеристик БСУ, поэтому в дальнейшем основное внимание уделяется данному классу случайных процессов.

Для описания поведения БСУ в данном классе марковских процессов необходимо: 1) ввести понятие состояния системы; 2) указать все состояния, в которых может находиться система; 3) составить граф состояний, т. е. указать пути возможных непосредственных переходов системы из состояния в состояние; 4) указать, в каком состоянии находится система в начальный момент времени, или задать распределение начальных состояний; 5) для каждого возможного перехода указать соответствующую интенсивность $\lambda_{ij}(t)$ потока событий, переводящих систему из состояния z_i в состояние z_j , где

$$\lambda_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(t, t + \Delta t)}{\Delta t}. \quad (4-1)$$

В дальнейшем рассматриваются только случаи, когда $\lambda_{ij}(t) = \lambda_{ij}$, т. е. интенсивности переходов не зависят от времени, что справедливо для однородных марковских процессов.

Остановимся кратко на вопросе, как вводится понятие состояния при описании процесса функционирования в классе марковских процессов. Определение состояния зависит от того, какие свойства в поведении системы нас интересуют. Например, при исследовании надежности функционирования БСУ с позиций безотказности и ремонтпригодности динамика ее представляет собой чередующийся процесс отказов и восстановлений как элементов, так и системы в целом. При исследовании производительности большой информационно-справочной системы ее динамика представляет собой процесс поступления в случайные моменты требований в систему и их обслуживание.

В первом случае состояние системы может быть определено как совокупность состояний элементов, каждый из которых может находиться либо в состоянии работоспособности, либо в состоянии отказа. Во втором — состояние системы определяется как число требований, находящихся в системе. Если требования однородны и обслуживание однофазное, то состояние определяется значением неотрицательной скалярной величины; если же требования неоднородны или обслуживание в системе многофазное, то состояние может быть определено как вектор числа требований, каждый компонент которого указывает либо число требований данного типа в системе, либо число требований, находящихся в определенной фазе системы.

Если определено полное множество состояний марковского процесса, то исчерпывающей характеристикой марковского процесса является совокупность вероятностей $P_j(t)$ того, что процесс в момент времени t будет находиться в состоянии z_j ; $j = 0, 1, 2, \dots, n$. Рассмотрим метод определения вероятностей $P_j(t)$.

Так как марковский процесс не обладает последствием, то справедлива следующая система уравнений:

$$P_j(t + \Delta t) = \sum_{i=1}^n P_i(t) P_{ij}(\Delta t); \quad j = 0, 1, 2, \dots, n. \quad (4-2)$$

Из условия $\sum_{i=1}^n P_{ji}(\Delta t) = 1$ следует, что

$$P_{jj}(\Delta t) = 1 - \sum_{i=1}^n P_{ji}(\Delta t); \quad i \neq j. \quad (4-3)$$

Подставляя (4-3) в систему уравнений (4-2), имеем:

$$P_j(t + \Delta t) = P_j(t) \left[1 - \sum_{\substack{i=1, \\ i \neq j}}^n P_{ji}(\Delta t) \right] + \sum_{\substack{i=1, \\ i \neq j}}^n P_i(t) P_{ij}(\Delta t); \quad j=0, 1, 2, \dots, n. \quad (4-4)$$

Так как в соответствии с (4-1)

$$P_{ij}(\Delta t) = \lambda_{ij} \Delta t + o(\Delta t),$$

где $o(\Delta t)$ — слагаемое, учитывающее конечность приращения Δt и имеющее порядок малости, больший чем Δt , то из системы уравнений (4-4) получаем:

$$\frac{P_j(t + \Delta t) - P_j(t)}{\Delta t} = - \left(\sum_{\substack{i=1, \\ i \neq j}}^n \lambda_{ji} \right) P_j(t) + \sum_{\substack{i=1, \\ i \neq j}}^n P_i(t) \lambda_{ij} + \frac{o(\Delta t)}{\Delta t}.$$

Совершая предельный переход при $\Delta t \rightarrow 0$, имеем следующую систему дифференциальных уравнений:

$$P'_j(t) = - \left(\sum_{\substack{i=1, \\ i \neq j}}^n \lambda_{ji} \right) P_j(t) + \sum_{\substack{i=1, \\ i \neq j}}^n P_i(t) \lambda_{ij}; \quad j=0, 1, 2, \dots, n; \quad (4-5)$$

$$P_0(0) = 1; \quad P_j(0) = 0.$$

Получили эквивалентное отображение динамики БСУ на языке марковских случайных процессов. Аналогичную систему дифференциальных уравнений можно построить формально по графу состояний, если придерживаться следующего правила.

В левой части каждого уравнения должна стоять производная вероятности состояния, а правая часть должна содержать столько членов, сколько стрелок связано с данным состоянием. Если стрелка направлена из состояния, соответствующий член имеет знак минус; если — в состояние, — знак плюс. Каждый член равен произведению интенсивности перехода, соответствующей данной стрелке, и вероятности того состояния, из которого исходит стрелка.

В результате решения системы дифференциальных уравнений определяются интересные нас характеристики

марковского случайного процесса вероятности $P_j(t)$. При решении данной системы дифференциальных уравнений необходимо учитывать, что

$$\sum_{i=1}^n P_j(t) = 1. \quad (4-6)$$

Если число состояний конечно и из каждого состояния графа можно перейти за то или иное число шагов в любое другое, то существуют предельные вероятности состояний

$$P_j = \lim_{t \rightarrow \infty} P_j(t); \quad j=0, 1, 2, \dots, n,$$

причем их значения не зависят от начального состояния системы.

На рис. 4-1, а показан граф состояний, удовлетворяющий поставленному условию; напротив, граф состояний, изображенный на рис. 4-1, б, не удовлетворяет данному условию.

При существовании предельного стационарного режима система случайным образом меняет свои состояния, но вероятность каждого из них уже не зависит от времени: каждое из состояний осуществляется с некоторой постоянной вероятностью, представляющей собой не что иное, как среднее относительное время пребывания системы в данном состоянии.

Для определения вероятностей P_j в системе уравнений (4-5) нужно положить все левые части уравнений равными нулю. Система дифференциальных уравнений (4-5) переходит в систему линейных алгебраических уравнений,

решение которой с учетом условия $\sum_{j=1}^n P_j = 1$ позволяет вычислить все предельные вероятности P_j . Рассмотрим применение изложенного выше подхода на следующем примере.

Пример 4-1. Исследуем влияние параметров программ на характеристики обмена в многопроцессорных вычислительных комплексах (ВК) с общей магистралью.

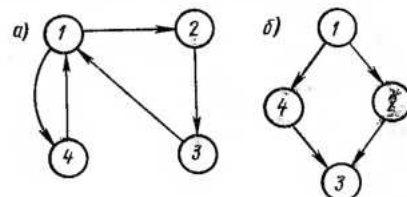


Рис. 4-1. Граф состояния системы: а — в случае наличия стационарного режима; б — в случае отсутствия стационарного режима

Структура двухпроцессорного ВК приведена на рис. 4-2, где *Пр* — процессоры, *ПЗУ* — постоянное запоминающее устройство, *ОЗУ* — оперативное запоминающее устройство.

В случае совместной работы двух процессоров при одной магистрали обмена командной и числовой информацией между процессорами, *ОЗУ* и *ПЗУ* будут, естественно, возникать конфликтные ситуации, когда оба процессора нуждаются в проведении обмена (ввод или вывод соответствующей информации). Если разрешение данного конфликта производится в пользу процессора, первого обратившегося к магистрали, то второй процессор будет находиться в состоянии ожидания ее освобождения. Причем время вынужденных простоев процессоров будет зависеть от временных характеристик программ. Чем больше длительность фазы автономной работы процессоров по сравнению с длительностью обмена в магистрали, тем меньше сказываются простои процессоров, тем выше производительность всего ВК в целом.

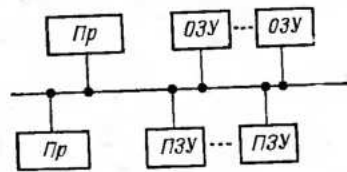


Рис. 4-2. Структура двухпроцессорного вычислительного комплекса

Задачей системотехнического анализа является исследование зависимости коэффициента загрузки магистрали η и коэффициента удлинения программ ϵ для данной структуры ВК от временных характеристик программ и пропускной способности магистрали. Динамика системы отражается в изменении состояния процессоров ВК. Каждый из процессоров может находиться в трех состояниях: 0 — работа, 1 — обмен, 2 — ожидание обмена. В соответствии с этим можно определить следующую совокупность состояний системы: (0; 0) — оба процессора работают; (1; 0) — первый процессор в состоянии обмена, второй в состоянии работы; (0; 1) — первый процессор в состоянии работы, второй — в состоянии обмена; (1; 2) — первый процессор в состоянии обмена, второй — в ожидании обмена; (2; 1) — первый процессор в состоянии ожидания обмена, второй — в состоянии обмена.

Граф, отражающий динамику системы, представлен на рис. 4-3. Для дальнейшего рассмотрения сделаем допущение о том, что длительность автономной фазы работы процессоров и длительность фазы их обмена являются случайными величинами, распределенными по показательному закону с параметрами λ_i , μ_i ($i = 1; 2$):

$$P(\tau_1^{(i)} \leq t) = 1 - e^{-\lambda_i t}; \quad P(\tau_2^{(i)} \leq t) = 1 - e^{-\mu_i t},$$

где $\tau_1^{(i)}$ — длительность фазы автономной работы i -го процессора; $\tau_2^{(i)}$ — длительность фазы обмена i -го процессора с памятью ($i = 1; 2$).

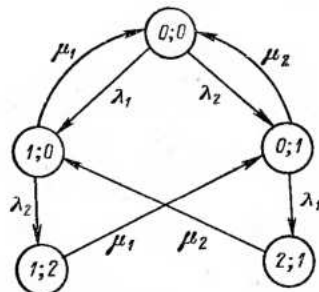


Рис. 4-3. Граф состояний для описания динамики двухпроцессорного вычислительного комплекса

В этом случае можно показать, что случайный процесс будет марковским. Соответствующие интенсивности переходов системы из состояния в состояние, определенные согласно (4-1), указаны на рис. 4-3. Как видно из графа поведения системы, стационарный режим существует и соответствующая система уравнений для него выглядит следующим образом:

$$\left. \begin{aligned} -(\lambda_1 + \lambda_2) P_{0;0} + \mu_1 P_{1;0} + \mu_2 P_{0;1} &= 0; \\ -(\lambda_2 + \mu_1) P_{1;0} + \lambda_1 P_{0;0} + \mu_2 P_{2;1} &= 0; \\ -(\lambda_1 + \mu_2) P_{0;1} + \lambda_2 P_{0;0} + \mu_1 P_{1;2} &= 0; \\ \lambda_2 P_{1;0} - \mu_1 P_{1;2} &= 0; \\ \lambda_1 P_{0;1} - \mu_2 P_{2;1} &= 0. \end{aligned} \right\}$$

Решая эту систему с учетом нормирующего условия (4-6), получим следующие выражения для вероятностей $P_{0;0}$, $P_{1;2}$, $P_{2;1}$:

$$P_{0;0} = \frac{1 + \alpha_1 \mu_1 / \mu_2 + \alpha_2 \mu_2 / \mu_1}{(1 + \alpha_1 \mu_1 / \mu_2 + \alpha_2 \mu_2 / \mu_1)(1 + \alpha_1 + \alpha_2 + \alpha_1 \alpha_2) + \alpha_1 \alpha_2 (1 + \alpha_1 + \alpha_2)};$$

$$P_{1;2} = \frac{\alpha_1 \alpha_2 \frac{\mu_2}{\mu_1} (1 + \alpha_1 \mu_1 / \mu_2 + \alpha_2)}{(1 + \alpha_1 \mu_1 / \mu_2 + \alpha_2 \mu_2 / \mu_1)(1 + \alpha_1 + \alpha_2 + \alpha_1 \alpha_2) + (1 + \alpha_1 + \alpha_2) \alpha_1 \alpha_2};$$

$$P_{2;1} = \frac{\alpha_1 \alpha_2 \frac{\mu_1}{\mu_2} (1 + \alpha_2 \mu_2 / \mu_1 + \alpha_1)}{(1 + \alpha_1 \mu_1 / \mu_2 + \alpha_2 \mu_2 / \mu_1)(1 + \alpha_1 + \alpha_2 + \alpha_1 \alpha_2) + \alpha_1 \alpha_2 (1 + \alpha_1 + \alpha_2)}.$$

Здесь $\alpha_1 = \lambda_1 / \mu_1$; $\alpha_2 = \lambda_2 / \mu_2$. Полученные выражения позволяют определить интересующие нас характеристики системы.

Коэффициент загрузки магистрали определяется следующим образом:

$$\eta = 1 - P_{0;0}.$$

Коэффициент удлинения первой программы

$$\epsilon_1 = \frac{T}{T - TP_{2;1}} = \frac{1}{1 - P_{2;1}},$$

а коэффициент удлинения второй программы

$$\epsilon_2 = \frac{T}{T - TP_{1;2}} = \frac{1}{1 - P_{1;2}}.$$

Данные выражения позволяют построить зависимости $\eta = f(\lambda_1, \lambda_2, \mu_1, \mu_2)$, $\epsilon_1 = f_2(\lambda_1, \lambda_2, \mu_1, \mu_2)$, $\epsilon_2 = f_3(\lambda_1, \lambda_2, \mu_1, \mu_2)$ и сделать вывод об эффективности использования указанной структуры ВК для решения выбранных задач обработки информации.

Следует отметить, что обычно установившийся режим поведения системы, а следовательно, и предельные вероятности состояний P_j для описывающего систему марковского процесса интересуют значительно чаще, чем переходный

режим, требующий определения вероятностей как функций времени. Аналитическое исследование предельных режимов поведения БСУ, подобно тому, как это было сделано в примере 4-1, возможно при малом числе состояний. При большом числе состояний модели возможно только численное исследование поведения системы с использованием ЭВМ. Из численных методов решения систем линейных алгебраических уравнений наиболее предпочтительным является метод использования алгоритма Гаусса для определения стационарных вероятностей [55]. Полученные оценки вероятностей состояний P_j используются в дальнейшем для получения интегральных значений показателей эффективности, усредненных по множеству состояний системы.

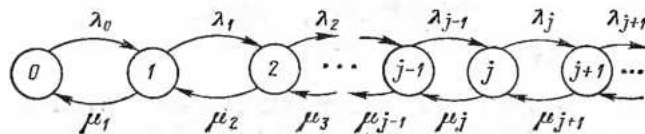


Рис. 4-4. Граф состояний процесса типа «гибель и размножение»

Далее рассмотрим широко распространенную частную модель марковских случайных процессов с дискретным числом состояний и непрерывным временем, используемую при анализе характеристик БСУ, — процесс типа «гибель и размножение». Происхождение данного термина ведет начало от биологических задач, где подобной схемой описывался процесс изменения численности биологической популяции.

Граф поведения марковского процесса такого вида представляет собой цепь состояний, в которой каждое из средних состояний ($z_1, z_2, \dots, z_{n-1}$) связано прямой и обратной связями с каждым из соседних состояний, а состояния z_0, z_n — только с одним соседним состоянием. Последнее справедливо при конечном значении n . Если n бесконечно, то цепь может быть продолжена на неограниченное число звеньев в направлении увеличения n (рис. 4-4).

Для такого процесса принята следующая система постулатов [20]:

- 1) $P_{j,j+1}(\Delta t) = \lambda_j \Delta t + o(\Delta t), \quad j \geq 0;$
- 2) $P_{j,j-1}(\Delta t) = \mu_j \Delta t + o(\Delta t), \quad j \geq 1;$
- 3) $P_{j,j}(\Delta t) = 1 - (\lambda_j + \mu_j) \Delta t + o(\Delta t), \quad j \geq 0;$

4) $\mu_0 = 0, \lambda_0 > 0; \lambda_j > 0, \mu_j > 0, j = 1, 2, \dots;$ если n конечно, то $\lambda_n = 0;$

5) $P_{ji}(\Delta t) = o(\Delta t)$, где $i \neq (j+1), i \neq (j-1), i \neq j$, а $o(\Delta t)$ — величина, имеющая порядок малости, больший чем Δt ; 6) интенсивности λ_j, μ_j зависят только от номера состояний j и не зависят от t .

Общая система дифференциальных уравнений, описывающая динамику такого процесса, будет выглядеть следующим образом:

$$\left. \begin{aligned} P'_0(t) &= -\lambda_0 P_0(t) + \mu_1 P_1(t), \quad j=0; \\ P'_j(t) &= -(\lambda_j + \mu_j) P_j(t) + \lambda_{j-1} P_{j-1}(t) + \\ &\quad + \mu_{j+1} P_{j+1}(t), \quad j \geq 1; \\ P_0(0) &= 1; \quad P_0(j) = 0, \quad j = 1, 2, \dots \end{aligned} \right\} \quad (4-7)$$

Вопросы выяснения существования и единственности решения полученной системы дифференциальных уравнений весьма подробно рассмотрены в [12]. Для того чтобы найденные вероятности $P_j(t)$ удовлетворяли условию $\sum_{j=1}^{\infty} P_j(t) = 1$, достаточно иметь ряд

$$\sum_{j=1}^{\infty} \prod_{k=1}^j \frac{\mu_k}{\lambda_k}$$

расходящимся.

Если при этом оказывается, что ряд $\sum_{j=1}^{\infty} \prod_{k=1}^j \frac{\lambda_{k-1}}{\mu_k}$ сходится, то существуют пределы вероятностей $P_j = \lim_{t \rightarrow \infty} P_j(t)$ ($j = 0, 1, 2, \dots$). Последнее условие, в частности, выполняется во всех случаях, когда начиная с некоторого k выполнено неравенство

$$\lambda_k / \mu_{k+1} < 1.$$

Как правило, в большинстве практических случаев это справедливо; следовательно, можно говорить о существовании предельных вероятностей состояний для процесса «гибель и размножение». Так же, как и в общем случае, будем интересоваться в дальнейшем только установившимся режимом. Для установившегося режима имеем сле-

дующую систему алгебраических уравнений:

$$\left. \begin{aligned} -\lambda_0 P_0 + \mu_1 P_1 &= 0; \\ -(\lambda_j + \mu_j) P_j + \lambda_{j-1} P_{j-1} + \mu_{j+1} P_{j+1} &= 0, \quad j \geq 1. \end{aligned} \right\} \quad (4-8)$$

Введем обозначение:

$$S_j = -\lambda_j P_j + \mu_{j+1} P_{j+1}.$$

Тогда уравнения (4-8) примут следующий вид:

$$S_0 = 0; \quad S_j - S_{j-1} = 0, \quad j \geq 1.$$

Отсюда следует, что при всех $j \geq 0$ $S_j = 0$. Из (4-8) находим, что

$$P_j = \frac{\lambda_{j-1}}{\mu_j} P_{j-1} = \prod_{k=1}^j \frac{\lambda_{k-1}}{\mu_k} P_0.$$

Из условия нормирования $\sum_{j=0}^{\infty} P_j = 1$ определяем

$$P_0 = \left[1 + \sum_{j=1}^{\infty} \prod_{k=1}^j \frac{\lambda_{k-1}}{\mu_k} \right]^{-1}.$$

Таким образом, предельные вероятности определяются следующим образом:

$$\left. \begin{aligned} P_0 &= \left[1 + \sum_{j=1}^{\infty} \prod_{k=1}^j \frac{\lambda_{k-1}}{\mu_k} \right]^{-1}; \\ P_j &= P_0 \prod_{k=1}^j \frac{\lambda_{k-1}}{\mu_k}. \end{aligned} \right\} \quad (4-9)$$

Применение марковского процесса указанного вида при анализе характеристик БСУ в основном ориентировано на два класса задач: исследование надежностных характеристик поведения систем с учетом отказов и восстановлений при отсутствии и наличии избыточности; исследование временных показателей поведения систем, отражающих производительность, пропускную способность, реакцию системы и т. д. В дальнейшем более детально рассматривается второй класс задач. При исследовании различных временных показателей обычно проводится формализация функционирования БСУ как систем массового обслуживания (СМО), по-

ведение которых при определенных условиях может быть описано на языке теории марковских процессов, в том числе и процессов типа «гибель и размножение».

4-2. ПРЕДСТАВЛЕНИЕ БСУ В ВИДЕ МОДЕЛЕЙ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ И ИХ КЛАССИФИКАЦИЯ

Все модели функционирования БСУ, которые можно рассматривать как СМО, имеют одну и ту же структуру (рис. 4-5).

Введем основные определения.

1. Последовательность событий будем называть **потоком**. Поток, состоящий из требований на обслуживание, назовем **поток требований**.

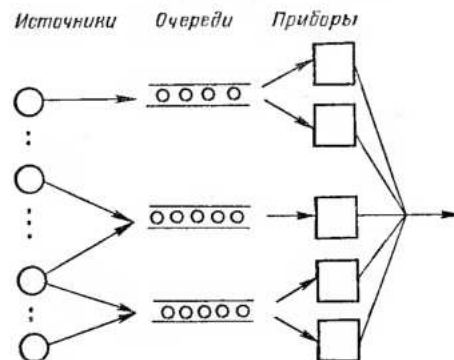


Рис. 4-5. Обобщенная структура модели БСУ как системы массового обслуживания

2. Поток требований, поступающих в обслуживающую систему, называется **входящим потоком**; поток требований, покидающих обслуживающую систему, называется **выходящим потоком**.

3. Первопричину возникновения требований независимо от их физической природы будем называть **источником**.

4. Совокупность очередей и приборов обслуживания будем называть **обслуживающей системой**. Каждое требование должно поступить на один из обслуживающих приборов, чтобы пройти соответствующее обслуживание, т. е. подвергнуться некоторой операции. Так как

в общем случае продолжительность обслуживания является случайной величиной, то требованиям приходится ожидать, пока не освободится обслуживающий прибор. В этом случае они находятся в накопителе, образуя одну или несколько очередей.

5. Алгоритм постановки требований в очередь будем называть *правилом формирования очереди*, а алгоритм взаимодействия обслуживающих приборов с очередью — *правилом обслуживания* или *дисциплиной обслуживания*.

Примерами БСУ, формализуемых в виде СМО, являются: система типа «Сирена» резервирования билетов на самолеты (требования — вызовы, справки о билетах; обслуживающая система — центр обработки данных, включающий многоаппаратный вычислительный комплекс); система обработки данных контроля сложных объектов (требования — события, связанные с необходимостью контроля параметров в определенные моменты времени; обслуживающая система — система коммутаторов, аналого-цифровых преобразователей, средств обработки и вывода информации о результатах контроля).

Существует достаточно большое число различных моделей СМО. Рассмотрим краткую классификацию основных моделей [43].

1. По *характеру источника требований* различают: источники с конечным числом требований, источники с бесконечным числом требований.

В первом случае в системе циркулирует конечное, и обычно постоянное, число требований. Требования после завершения обслуживания возвращаются в источник, где они пребывают в течение некоторого времени; затем вновь поступают в систему. Во втором случае источник генерирует бесконечное число требований и работа источника никак не зависит от работы обслуживающей системы. Системы с конечным числом требований называются *замкнутыми*; системы с бесконечным числом требований — *разомкнутыми*.

2. По *отсутствию или наличию возможности ожидания* для требования выделяют: системы с *отказами* (в таких системах требование, поступившее в момент, когда все обслуживающие приборы заняты, получает отказ, покидает систему и в дальнейшем процессе обслуживания не участвует); системы с *ожиданием* (в таких системах требование, поступившее в момент, когда все приборы

обслуживания заняты, становится в очередь и ожидает, пока не освободится один из приборов обслуживающей системы). Как только освобождается прибор, принимается к обслуживанию одно из требований, стоящих в очереди.

Системы с ожиданием делятся на системы с *неограниченным ожиданием* и системы с *ограниченным ожиданием*. В случае неограниченного ожидания любое требование, рано или поздно пришедшее в систему, будет обслужено. В системах с ограниченным ожиданием накладываются те или иные ограничения: на длину очереди; на время пребывания требования в очереди, после которого требование покидает очередь и уходит; на общее время пребывания требования в системе.

3. По *числу приборов* в обслуживающей системе различают: *одноканальные* системы массового обслуживания (один обслуживающий прибор); *многоканальные* системы массового обслуживания (несколько обслуживающих приборов).

4. По *числу этапов (фаз)* обслуживания в обслуживающей системе различают: *однофазные* и *многофазные* системы.

5. По *правилу формирования очереди* различают: системы массового обслуживания с *общей очередью* (общий накопитель); системы массового обслуживания с *несколькими очередями* (накопитель, разделенный на зоны).

Для общей очереди ограниченной длины возможны три основных алгоритма постановки требований в очередь: а) алгоритм постановки требований в очередь по «кольцу» — требования ставятся в очередь «по кольцу». В этом случае требования заносятся в очередь последовательно в порядке поступления, а при достижении ограничения запись продолжается с первой позиции; при этом стираются требования, оказавшиеся на месте, куда должно быть поставлено требование; б) заполнение очереди в порядке поступления требований с поиском свободных мест, при отсутствии свободных мест в очереди требование получает отказ; в) приоритетные дисциплины (правила) формирования очередей основаны на том, что требования неоднородны по значимости, поэтому при поступлении каждого нового требования производится поиск свободного места в очереди и, если его не оказывается, то производится поиск самого старого требования, имеющего приоритет, равный или более низкий,

чем приоритет поступившего требования; если такое требование найдено, то оно стирается, а на его место записывается поступившее. Если такое требование не найдено, то поступившее требование получает отказ.

В случае существования нескольких очередей постановка однотипных требований в каждую очередь может осуществляться либо в порядке поступления, либо «по кольцу».

6. По правилу (дисциплине) обслуживания различают системы массового обслуживания: с бесприоритетными дисциплинами обслуживания, из которых наиболее распространенными являются правила: «первым пришел — первым обслужен»; циклическое обслуживание по одному требованию из различных источников; «последним пришел — первым обслужен»; с приоритетными дисциплинами, т. е. требование выбирается на обслуживание в соответствии с номером приоритета. При этом, если в течение времени обслуживания приходит требование с более высоким приоритетом, то обслуживание может быть прервано (абсолютный приоритет) либо не прервано до полного завершения, после чего на обслуживание принимается требование с более высоким приоритетом (относительный приоритет).

Рассмотренная классификация СМО является далеко не полной и охватывает только основные их виды, используемые в качестве наиболее распространенных моделей БСУ.

Определить модель системы в виде СМО — значит, задать характеристики всех основных ее элементов: характеристики потока требований и источника требований; правило формирования очереди; характеристики обслуживающей системы (характеристики закона обслуживания, число каналов, число фаз обслуживания, правило обслуживания).

В зависимости от вида и назначения системы при оценке ее эффективности могут применяться те или иные показатели. Для систем, представимых в виде моделей СМО с отказами, одной из важнейших характеристик их эффективности является либо абсолютная пропускная способность (среднее число требований, которое может обслужить система за единицу времени), либо относительная пропускная способность (средняя доля поступивших требований, обслуживаемая системой в единицу времени). При этом могут интересовать и другие косвенные характеристики: среднее число занятых каналов, среднее относительное время простоя системы в целом и отдельного канала и др.

Для систем без потерь (с неограниченным ожиданием) весьма важными показателями эффективности качества обслуживания являются: среднее число требований в очереди, среднее число требований в системе, среднее время ожидания требования в очереди, среднее время пребывания требования в системе, коэффициент простоя или коэффициент загрузки обслуживающей системы и т. д. Поэтому основной задачей анализа при использовании моделей СМО является отыскание функциональных зависимостей выбранных показателей эффективности от характеристик входящего потока, параметров, характеризующих обслуживаемую систему, от правила формирования очереди и дисциплины обслуживания.

Рассмотрение всех возможных аналитических методов исследования различных перечисленных выше моделей систем массового обслуживания при всем разнообразии перечисленных параметров является весьма сложной задачей. Современная теория массового обслуживания обладает весьма интересными аналитическими подходами для анализа характеристик как марковских, так и немарковских систем массового обслуживания, динамика которых не описывается в классе марковских процессов [43]. В дальнейшем остановимся только на моделях систем массового обслуживания, описываемых в классе марковских процессов типа «гибель и размножение» как наиболее распространенных на практике при выполнении инженерных расчетов показателей эффективности.

Функционирование системы массового обслуживания всегда описывается процессом типа «гибель и размножение», если источник содержит только однородные требования, поток требований является простейшим (стационарным пуассоновским потоком), закон обслуживания является показательным, дисциплина обслуживания: «первым пришел — первым обслужен».

Для простейшего потока число событий k , происшедших на интервале длиной T , подчиняется пуассоновскому распределению

$$P_k(T) = (\lambda T)^k e^{-\lambda T} / k!, \quad k = 0, 1, 2, \dots,$$

где λ — интенсивность потока, определяемая как

$$\lambda = \lim_{\Delta t \rightarrow 0} \frac{\sum_{k=1}^{\infty} k P_k(\Delta t)}{\Delta t},$$

а длительность интервала между требованиями в потоке распределена с плотностью распределения

$$f(\tau) = \lambda e^{-\lambda\tau}, \tau \geq 0; \quad f(\tau) = 0, \tau < 0,$$

которую обычно называют показательным законом.

Аналогичный закон распределения должен быть и для длительности обслуживания, только с другой интенсивностью μ — интенсивностью обслуживания. Интенсивность потока и интенсивность обслуживания экспериментально обычно определяются через среднюю длину интервала между требованиями $\bar{t}$ и среднюю длительность обслуживания $\bar{t}_0$:

$$\lambda = 1/\bar{t}; \quad \mu = \frac{1}{\bar{t}_0}.$$

Для входного потока требований в БСУ могут быть построены и другие модели.

Поток требований может быть регулярным, если требования следуют одно за другим через строго определенные промежутки времени. Обобщением простейшего потока являются потоки Пальма (потоки с ограниченным последствием), если промежутки времени между последовательными событиями представляют собой независимые одинаково распределенные случайные величины. Примером потоков Пальма являются так называемые потоки Эрланга, получаемые в результате «просеивания» простейших потоков. Поток Эрланга r -го порядка называется поток, получающийся, если в простейшем потоке сохранить каждое r -е требование, а остальные не принимать во внимание. Плотность распределения длительности интервала между событиями в потоке Эрланга r -го порядка выглядит следующим образом:

$$f_r(\tau) = \frac{\lambda (\lambda\tau)^{r-1}}{(r-1)!} e^{-\lambda\tau}, \quad \tau \geq 0; \quad f(\tau) = 0, \tau < 0.$$

При $r = 1$ получаем простейший поток. Как показано в [8], при $r \rightarrow \infty$ поток Эрланга заданной интенсивности приближается к регулярному потоку с постоянным интервалом между событиями. Это свойство потоков Эрланга удобно в практических применениях. Оно дает возможность, задаваясь различными значениями r , получать потоки, обладающие различным последствием — от полного отсут-

ствия последствия ($r = 1$) до жесткой функциональной связи между моментами появления событий ($r = \infty$). Аналогично и длительность обслуживания может быть либо постоянной величиной, либо иметь закон распределения Эрланга и др.

Поток может быть многомерным, т. е. от нескольких источников требований. Для многомерного входного потока при однородных бесконечных источниках возможно его эквивалентирование и представление в виде одного источника с общим суммарным входным потоком. Если потоки неоднородные, то такая эквивалентная замена недопустима. Так, например, при анализе УВМ как системы массового обслуживания обычно выделяют четыре основных типа входных потоков: 1) сообщения внешних абонентов; 2) сообщения, сформированные при решении основных функциональных задач; 3) требования на включение периодических программ; 4) требования на включение подпрограмм, выполняющих анализ сигналов аппаратного контроля УВМ и внешних специализированных устройств.

Неоднородность предполагает различие как в виде поступающей информации, так и в характеристиках потока и характеристиках обслуживания требований из разных потоков. Принимая во внимание физический характер формирования потоков, можно определить закон распределения длительности интервалов между требованиями потока. Так, например, требования на включение периодических программ при работе УВМ могут образовать регулярный поток, а требования на включение подпрограмм, выполняющих анализ сигналов аппаратного контроля УВМ и внешних специализированных устройств, составляют поток, близкий к пуассоновскому. При многомерном потоке, принимая во внимание различную важность информации, содержащейся в требованиях различных потоков, вводят систему приоритетов, которая в дальнейшем учитывается при обслуживании.

Однако во многих из этих случаев поведение СМО не описывается в классе марковских процессов и для исследования таких моделей обычно используют либо аналитические методы, основанные на системе специальных преобразований для сведения немарковских процессов к марковским, либо методы статистического моделирования.

4-3. МОДЕЛИ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ, ОПИСЫВАЕМЫЕ ПРОЦЕССАМИ ТИПА «ГИБЕЛЬ И РАЗМНОЖЕНИЕ», И ИХ ХАРАКТЕРИСТИКИ

Марковский процесс типа «гибель и размножение» в установившемся режиме характеризуется предельными вероятностями, определяемыми по формулам:

$$P_0 = \left[1 + \sum_{j=1}^{\infty} \prod_{k=1}^j \frac{\lambda_{k-1}}{\mu_k} \right]^{-1}; \quad P_j = P_0 \prod_{k=1}^j \frac{\lambda_{k-1}}{\mu_k}.$$

Рассмотрим ряд моделей СМО при указанных выше допущениях и определим основные их характеристики.

Многоканальные и одноканальные системы без потерь с неограниченным ожиданием и источником с бесконечным числом требований. В этом случае предполагается, что система содержит N обслуживающих одинаковых приборов с одинаковой интенсивностью обслуживания μ при входном потоке с интенсивностью λ .

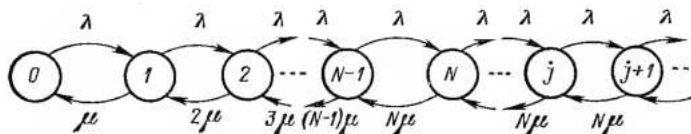


Рис. 4-6. Граф динамики многоканальных систем без потерь с неограниченным ожиданием и источником бесконечных требований

Граф динамики подобной системы представлен на рис. 4-6, где j — число требований в системе; $\lambda_j = \lambda$, $\mu_j = j\mu$ при $1 \leq j \leq N$; $\mu_j = N\mu$ при $j > N$, $j = 0, 1, 2, \dots$. На основании выражений (4-9) получим

$$\left. \begin{aligned} P_j &= (\lambda/\mu)^j \frac{1}{j!} P_0, & 1 \leq j \leq N; \\ P_j &= (\lambda/\mu)^j \frac{1}{N! N^{j-N}} P_0, & j > N; \\ P_0 &= \left[1 + \sum_{j=1}^N (\lambda/\mu)^j \frac{1}{j!} + \sum_{j=N+1}^{\infty} (\lambda/\mu)^j \frac{1}{N! N^{j-N}} \right]^{-1} = \\ &= \left[1 + \sum_{j=1}^N (\lambda/\mu)^j \frac{1}{j!} + (\lambda/\mu)^{N+1} \frac{\mu}{N! (N\mu - \lambda)} \right]^{-1}. \end{aligned} \right\} \quad (4-10)$$

Граф динамики одноканальной СМО представлен на рис. 4-7, а соответствующие выражения для определения вероятностей P_j могут быть получены из (4-10) как частный случай при $N = 1$:

$$P_j = (\lambda/\mu)^j P_0, \quad j = 1, 2, \dots; \quad P_0 = \left[1 + \sum_{j=1}^{\infty} (\lambda/\mu)^j \right]^{-1}. \quad (4-11)$$

Из (4-10), (4-11) вытекает, что существование конечной средней очереди для многоканальных систем возможно при условии $\lambda < N\mu$ или $\rho_N < 1$, где $\rho_N = \lambda/(N\mu)$, а для одноканальных — при условии $\lambda < \mu$ или $\rho < 1$, где $\rho = \lambda/\mu$.

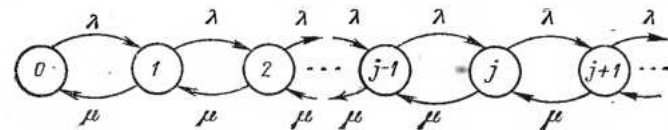


Рис. 4-7. Граф динамики одноканальных систем массового обслуживания без потерь с неограниченным ожиданием и источником бесконечных требований

При этих ограничениях могут быть получены все интересные показатели эффективности на основе формул (4-10), (4-11).

Вероятность простоя:
для многоканальных систем

$$P_0 = \left[1 + \sum_{j=1}^N \rho^j \frac{1}{j!} + \rho^{N+1} \frac{1}{N! (N - \rho)} \right]^{-1}; \quad (4-12)$$

для одноканальных систем

$$P_0 = \left[1 + \sum_{j=1}^{\infty} \rho^j \right]^{-1} = 1 - \rho, \quad (4-13)$$

где ρ определяется как коэффициент загрузки одного обслуживающего прибора.

Среднее число требований в очереди

$$\bar{n}_0 = \sum_{j=N+1}^{\infty} (j - N) P_j. \quad (4-14)$$

Подставляя в (4-14) выражения для P_j и проводя простейшие преобразования, для многоканальных систем по-

лучим:

$$\bar{n}_0 = \frac{\rho^{N+1} P_0}{N N! (1-\rho_\Sigma)^2}. \quad (4-15)$$

Соответственно для одноканальных систем

$$\bar{n}_0 = \frac{\rho^2}{1-\rho}. \quad (4-16)$$

Среднее число занятых каналов

$$\bar{N}_s = \lambda/\mu = \rho. \quad (4-17)$$

Среднее число требований в системе

$$J = \bar{n}_0 + \bar{N}_s. \quad (4-18)$$

Среднее время ожидания требования в очереди для многоканальных систем

$$t_{\text{ож}} = \sum_{j=N}^{\infty} \frac{1}{N\mu} (j+1-N) P_j = \frac{\rho^N P_0}{N\mu N! (1-\rho_\Sigma)^2}. \quad (4-19)$$

При $N = 1$ из (4-19) получаем выражение для $t_{\text{ож}}$ в одноканальной системе

$$t_{\text{ож}} = \frac{\rho}{\mu(1-\rho)}. \quad (4-20)$$

Среднее время пребывания требования в системе

$$t_c = t_{\text{ож}} + 1/\mu, \quad (4-21)$$

т. е. равно среднему времени ожидания плюс среднее время обслуживания.

Многоканальные и одноканальные системы с потерями и источником бесконечного числа требований. Предполагаем, что потери возникают по причине ограниченности очереди (накопителя). Считаем, что очередь ограничена по длине величиной m , т. е. имеются m мест для ожидания требований. В случае, если все приборы заняты и накопитель полностью заполнен, требование теряется. Граф динамики многоканальной системы остается таким же, как и на рис. 4-6, только число состояний в графе конечно и равно $N + m + 1$, включая нулевое состояние. При этом в системе может находиться только $N + m$ требований.

Из формул (4-10) могут быть получены все интересующие показатели эффективности для данного класса систем массового обслуживания.

Вероятность нахождения в системе j требований

$$\left. \begin{aligned} P_j &= \rho^j \frac{1}{j!} P_0, & 1 \leq j \leq N; \\ P_j &= \rho^j \frac{1}{N!} \frac{1}{N^{j-N}} P_0, & N < j \leq N+m; \\ P_0 &= \left[1 + \sum_{j=1}^N \rho^j \frac{1}{j!} + \sum_{j=N+1}^{N+m} \rho^j \frac{1}{N!} \frac{1}{N^{j-N}} \right]^{-1} = \\ &= \left[1 + \sum_{j=1}^{N-1} \rho^j \frac{1}{j!} + \frac{\rho^N (1-\rho_\Sigma^{m+1})}{N! (1-\rho_\Sigma)} \right]^{-1}. \end{aligned} \right\} \quad (4-22)$$

Вероятность потери требований определяется, исходя из того, что требования теряются, если многоканальная система находится в состоянии $N + m$, т. е. все приборы заняты и накопитель полностью заполнен. Поэтому вероятность потерь для многоканальной системы определяется вероятностью нахождения системы в $N + m$ состоянии, т. е.

$$P_{\text{пот}} = P_{N+m} = \rho^{N+m} \frac{1}{N!} \cdot \frac{1}{N^m} \cdot P_0. \quad (4-23)$$

Для одноканальных систем при $N = 1$ получаем из (4-23)

$$P_{\text{пот}} = P_{m+1} = \frac{\rho^{m+1} (1-\rho)}{1-\rho^{m+2}}. \quad (4-24)$$

Абсолютная пропускная способность: $A = \lambda (1 - P_{\text{пот}})$. Среднее число требований в очереди

$$\begin{aligned} \bar{n}_0 &= \sum_{j=N+1}^{N+m} (j-N) P_j = \\ &= \frac{\rho^{N+1}}{N N!} \left[\frac{1-\rho_\Sigma^m [(m+1)-m\rho_\Sigma]}{(1-\rho_\Sigma)^2} \right] P_0. \end{aligned} \quad (4-25)$$

Среднее число занятых каналов определяется как отношение абсолютной пропускной способности системы к интенсивности обслуживания одного канала:

$$\bar{N}_s = \frac{A}{\mu} = \frac{\lambda}{\mu} (1 - P_{\text{пот}}). \quad (4-26)$$

Среднее число требований в системе

$$J = \bar{n}_0 + \bar{N}_s. \quad (4-27)$$

Среднее время ожидания требования в очереди

$$\begin{aligned} t_{\text{ож}} &= \sum_{j=N}^{N+m} \frac{1}{N\mu} (j+1-N) P_j = \\ &= \frac{\rho^N P_0}{N\mu N!} \cdot \frac{1-(m+1)\rho_\Sigma^m + m\rho_\Sigma^{m+1}}{(1-\rho_\Sigma)^2}. \end{aligned} \quad (4-28)$$

Среднее время пребывания требования в системе

$$t_c = t_{\text{ож}} + 1/\mu. \quad (4-29)$$

Все интересующие показатели для одноканальных систем массового обслуживания могут быть получены из соответствующих формул при $N = 1$.

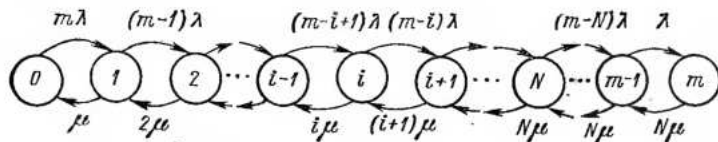


Рис. 4-8. Граф динамики многоканальных систем массового обслуживания без потерь с источником конечных требований

Многоканальные и одноканальные системы без потерь с источником конечного числа требований. Рассмотренные выше модели СМО характеризовались тем, что требования приходили в систему из источника с бесконечным числом требований и интенсивность потока требований не зависела от состояния системы. В данном случае рассматриваются системы массового обслуживания, в которых интенсивность потока поступающих требований зависит от состояния системы. В дальнейшем будем считать, что имеется N каналов обслуживания с интенсивностью обслуживания μ и источник, содержащий конечное число m требований, поступающих с интенсивностью λ , причем $m > N$. Граф состояний, отражающий динамику такой системы и вытекающий из общего графа процессов типа «гибель и размножение», представлен на рис. 4-8. В такой системе содержится всегда конечное число требований m , поэтому данный граф содержит всегда конечное число состояний, равное $m+1$. Интенсивность переходов в графе определяется следующим образом:

$$\left. \begin{aligned} \lambda_j &= (m-j)\lambda, & 0 \leq j \leq m-1; \\ \mu_j &= j\mu, & 1 \leq j \leq N; \\ \mu_j &= N\mu, & N < j \leq m. \end{aligned} \right\}$$

В соответствии с общими формулами (4-9) для вероятностей P_j можно, используя конкретные значения λ_j и μ_j , для данной модели СМО определить вероятности соответствующих состояний:

$$\left. \begin{aligned} P_j &= \left(\frac{\lambda}{\mu}\right)^j \frac{m!}{(m-j)! j!} P_0, & 1 \leq j \leq N; \\ P_j &= \left(\frac{\lambda}{\mu}\right)^j \frac{m!}{(m-j)! N! N^{j-N}} P_0, & N < j \leq m; \\ P_0 &= \left[1 + \sum_{j=1}^N \rho^j C_m^j + \sum_{j=N+1}^m \rho^j C_m^j \frac{j!}{N! N^{j-N}} \right]^{-1}, & \rho = \lambda/\mu. \end{aligned} \right\} \quad (4-30)$$

В соответствии с данными формулами могут быть определены все остальные показатели эффективности для данного класса моделей СМО.

Вероятность простоя обслуживающей системы

$$P_0 = \left[1 + \sum_{j=1}^N \rho^j C_m^j + \sum_{j=N+1}^m \rho^j C_m^j \frac{j!}{N! N^{j-N}} \right]^{-1}. \quad (4-31)$$

Среднее число требований в очереди

$$\bar{n}_0 = \sum_{j=N+1}^m (j-N) P_j = \sum_{j=N+1}^m (j-N) C_m^j \frac{\rho^j j!}{N! N^{j-N}} P_0. \quad (4-32)$$

Среднее число занятых обслуживающих каналов

$$\bar{N}_s = N - \sum_{j=0}^{N-1} (N-j) P_j. \quad (4-33)$$

Среднее число требований в системе

$$\bar{j} = \bar{n}_0 + \bar{N}_s. \quad (4-34)$$

Среднее время ожидания обслуживания и среднее время пребывания в системе получены в [24] и определяются следующими формулами:

$$t_{\text{ож}} = \frac{\bar{j}}{\lambda(m-j)} - \frac{1}{\mu}; \quad t_c = \frac{\bar{j}}{\lambda(m-j)}. \quad (4-35)$$

Выражения (4-35) получены из того условия, что в замкнутых системах средний входящий поток равен среднему выходящему. Полученные выше соотношения для многоканальных систем значительно упрощаются для случая $N = 1$ [30]:

Вероятность простоя обслуживающего прибора

$$P_0 = \left[\sum_{j=0}^m \frac{m!}{(m-j)!} \rho^j \right]^{-1}. \quad (4-36)$$

Среднее число требований в очереди

$$\bar{n}_0 = m - \frac{\mu + \lambda}{\lambda} (1 - P_0). \quad (4-37)$$

Среднее число требований в системе

$$\bar{j} = m - \frac{(1 - P_0) \mu}{\lambda}. \quad (4-38)$$

Среднее время ожидания в очереди

$$t_{\text{ож.}} = \frac{1}{\mu} \left[\frac{m}{1 - P_0} - \frac{\mu + \lambda}{\lambda} \right]. \quad (4-39)$$

Среднее время пребывания в системе

$$t_c = \frac{1}{\mu} \left[\frac{m}{1 - P_0} - \frac{\mu}{\lambda} \right]. \quad (4-40)$$

Рассмотренные выше модели СМО, описываемые в классе марковских процессов типа «гибель и размножение», не исчерпывают всех возможных приложений таких процессов в теории БСУ. Так, например, в [8] на основе рассмотренного аппарата рассматриваются многоканальные многофазные СМО с учетом взаимодействия каналов, ошибок при обслуживании и возможных повторений обслуживания отдельных требований и др.

Далее рассмотрим конкретные примеры применения процессов типа «гибель и размножение» для решения различных системотехнических задач.

Пример 4-2. Определим необходимую производительность периферийных технических средств для обслуживания оперативных требований, поступающих в информационно-справочную систему. Структурная схема системы представлена на рис. 1-4.

Из всего большого комплекса системотехнических задач, возникающих при проектировании такого класса систем, выделим только одну: определение необходимой производительности подсистемы периферийных технических средств (блок коммутации сообщений и пульта-мониторы, устанавливаемые у кассиров). Пульт-монитор в дальнейшем будем определять как терминал, предназначенный для ввода информации, регистрации и одновременного визуального вывода информации на пульт. Блок коммутации предназначен для концентрации потока требований, выбора свободного канала системы передачи данных и хранения требований в буферном запоминающем устройстве (в случае отсутствия свободного канала для передачи).

В качестве показателя для данного исследования целесообразно выбрать среднее время пребывания требования в рассматриваемой

подсистеме. Параметры рассматриваемой подсистемы: число пультов-мониторов m , число каналов передачи данных n , объем буферного запоминающего устройства k — должны быть выбраны таким образом, чтобы предотвратить образование бесконечной очереди на входе в рассматриваемую подсистему. При этом средняя производительность терминала и средняя производительность передачи требования по каналу связи считаются заданными.

В данной системе стохастический поток требований проходит последовательное обслуживание двумя группами приборов (терминалы и каналы системы передачи данных), по m и n приборов в каждой группе (соответственно $m \geq 1$, $n \geq 1$). Приборы первой группы характеризуются одинаковой средней производительностью μ_1 , средняя производительность каждого из приборов второй группы μ_2 . Перед первой группой приборов допускается неограниченная очередь требований входного потока. Требования, прошедшие обслуживание на приборах первой группы, поступают либо на один из свободных приборов второй группы, либо (при отсутствии таковых) в буферный накопитель емкостью в k заявок. Если требование в момент завершения обслуживания на приборе первой группы застаёт буферный накопитель заполненным, то оно не покидает указанного прибора до момента освобождения места в буферном накопителе. На все это время прибор первой группы оказывается заблокированным. Свободные места в буферном накопителе, так же, как и обслуживающие приборы, занимают в естественном порядке поступления требований. Предполагается, что входной поток — пуассоновский с интенсивностью λ , а длительности обслуживания распределены по показательному закону соответственно со средними значениями μ_1^{-1} , μ_2^{-1} . Построенная модель является многоканальной двухфазной СМО с неограниченной очередью и блокировкой приборов в первой фазе обслуживания и ограниченной очередью во второй фазе.

Аналитическое исследование построенной модели проведем в наихудшем режиме — режиме сильных нагрузок (пиковых нагрузок) [57]. Режим сильных нагрузок характеризуется тем, что на входе рассматриваемой подсистемы с вероятностью 1 всегда имеется ненулевая очередь. В этом случае интересующее нас среднее время пребывания требования в подсистеме от момента поступления на обслуживание в первую группу приборов до момента выхода из второй фазы рассчитывается по формуле

$$t_c = \mu_1^{-1} + \mu_2^{-1} + t_{\text{ож.}}, \quad (4-41)$$

где $t_{\text{ож.}}$ — среднее время ожидания обслуживания во второй фазе рассматриваемой подсистемы.

Значение $t_{\text{ож.}}$ определяется согласно выражению (4-19), в соответствии с которым

$$t_c = \mu_1^{-1} + \mu_2^{-1} + \frac{1}{n\mu_2} \sum_{j=n}^{n+k+m} (j-n+1) P_j,$$

где j — текущее число требований, находящихся во второй фазе обслуживания.

В процессе функционирования j может принимать следующие значения: $j = 0$ — все приборы первой фазы загружены, накопитель и все приборы второй фазы свободны; $0 < j \leq n$ — все приборы первой фазы заняты, накопитель пуст, а приборы второй фазы частично или полностью заняты; $n < j \leq n+k$ — все приборы первой и второй фаз

заняты, накопитель частично или полностью заполнен; $n + k < j \leq n + k + m$ — все приборы первой и второй фаз заняты, накопитель полностью занят, часть или все приборы первой фазы заблокированы, и требования, находящиеся в них, ждут освобождения места в накопителе.

Отсюда видно, что в режиме сильных нагрузок динамика подсистемы выражается только изменением числа требований, находящихся во второй фазе (на обслуживании, в накопителе либо в приборах первой фазы, находящихся в состоянии блокировки). Это позволяет свести исследование двухфазной СМО к исследованию однофазной СМО. Определив состояние системы как j , можно построить граф, отображающий ее динамику (рис. 4-9), структура которого соответствует графу рассмотренных уже марковских процессов типа «гибель и размножение».

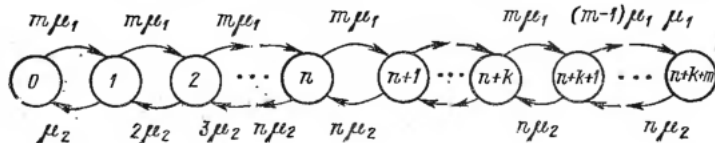


Рис. 4-9. Граф поведения подсистемы периферийных технических средств в режиме пиковых нагрузок

Соответствующая система дифференциально-разностных уравнений имеет вид:

$$\begin{aligned} P'_0(t) &= -m\mu_1 P_0(t) + \mu_2 P_1(t); \\ P'_j(t) &= -P_j(t)(\lambda_j + \mu_j) + \lambda_{j-1} P_{j-1}(t) + \mu_{j+1} P_{j+1}(t); \\ j &= 1, 2, \dots, n+k+m, \end{aligned}$$

где

$$\lambda_j = \begin{cases} m\mu_1, & 0 < j \leq n+k; \\ (m+n+k-j)\mu_1, & n+k < j \leq n+k+m; \end{cases}$$

$$\mu_j = \begin{cases} j\mu_2, & 0 < j \leq n; \\ n\mu_2, & n < j \leq n+k+m. \end{cases}$$

На основе общих выражений для стационарных вероятностей (4-9) имеем

$$P_j = \begin{cases} \frac{(m\rho)^j}{j!} P_0, & j = 1, 2, \dots, n, \rho = \mu_1/\mu_2; \\ \frac{(m\rho)^j}{n! n^{j-n}} P_0, & n+1 \leq j \leq n+k; \\ \frac{m^{n+k} m! \rho^j}{n^{j-n} n! (m+n+k-j)!} P_0, & n+k+1 \leq j \leq n+k+m; \end{cases}$$

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

На основе полученных вероятностей состояний можно определить значение $t_{\text{ож}}$. Эквивалентная средняя производительность технических средств подсистемы системы в режиме пиковых нагрузок определяется как $\mu_c = m t_c^{-1}$, откуда следует необходимое и достаточное условие наличия стационарной конечной средней очереди на входе $\lambda < m t_c^{-1}$.

Данное соотношение может быть исходным для выбора параметров m, n, k , характеризующих рассматриваемую подсистему.

При мер 4-3. Проанализируем эффективность организации мультипрограммной работы автоматизированной системы контроля (АСК) в условиях разделения времени.

Алгоритм контроля сложных объектов в целом можно представить в виде некоторой совокупности алгоритмов проверки отдельных независимых подсистем, последовательность выполнения которых в общем случае можно считать безразличной. Для выполнения данной совокупности алгоритмов в последовательном режиме используется: некоторый набор устройств, в который входят центральное устройство обработки информации (ЦУО), включающее ЭВМ; класс устройств, выполняющих различные операции по управлению, сбору, преобразованию, отображению информации, циркулирующей между контролируемым объектом и АСК. В дальнейшем будем называть такие устройства функциональными (ФУ). Любая программа контроля при реализации соответствующего алгоритма может быть представлена в виде последовательности работы операторов, характеризующих работу ЦУО и ФУ, причем фаза оператора ЦУО чередуется с фазой оператора, отражающего работу какого-либо ФУ.

Применение мультипрограммного режима при создании АСК оказывается целесообразным в силу того, что среднее время автономной фазы работы ФУ оказывается значительно больше, чем среднее время формирования команд управления и обработки информации в ЦУО. Это позволяет организовать совмещенный контроль при параллельном выполнении нескольких программ, в результате чего сокращается общее время функционирования АСК.

На этапе макроектирования АСК необходимо определить степень повышения производительности АСК при переходе с последовательного режима контроля на мультипрограммный в зависимости от числа параллельно выполняемых программ и выбранных дисциплин диспетчеризации. В качестве показателя эффективности введения мультипрограммной организации может быть выбрано отношение

$$v = T_{\text{п}}/T_{\text{мл}},$$

где $T_{\text{п}}$ — время проверки объекта при последовательном режиме; $T_{\text{мл}}$ — время проверки объекта в мультипрограммном режиме (совмещенная проверка).

Для анализа эффективности организации мультипрограммного режима в АСК может быть принято представление АСК в виде одноканальной замкнутой СМО.

В качестве m требований источника конечных требований рассматривается m совмещаемых программ, причем время пребывания i -го требования в источнике эквивалентно времени работы соответствующего ФУ по i -й программе контроля.

Требование на работу ЦУО по i -й программе возникает в момент окончания фазы работы ФУ по i -й программе (требование с номером i выходит из источника конечных требований), а требование на работу ФУ

возникает в момент окончания фазы работы ЦУО по i -й программе (требование с номером i возвращается в источник).

Статистический анализ временных характеристик программ показал, что длительность фаз работы ЦУО ($t_{ЦУО}$) и длительности фаз работы ФУ ($t_{ФУ}$) по каждой из программ можно приближенно рассматривать как случайные величины, подчиняющиеся показательному закону распределения с параметрами $\mu = 1/t_{ЦУО}$ и $\lambda = 1/t_{ФУ}$. Следовательно, все программы, а значит, и требования можно считать однородными по своим временным характеристикам.

Прибором обслуживания является ЦУО, переключающееся с программы на программу по мере выполнения своей работы, причем время перехода с одной программы на другую программу не учитываем. Считаем, что правило выбора программ на обслуживание в ЦУО соответствует дисциплине: «первым пришел — первым обслужен».

Все ФУ считаются совместимыми, т. е. допускается их параллельная работа; ФУ, работающие в реальном масштабе времени, не рассматриваются.

Рассмотренный выше показатель эффективности мультипрограммной организации, как показано в [19], может быть выражен через коэффициенты загрузки ЦУО при совмещении и последовательном контроле:

$$v = \eta_m / \eta_1,$$

где η_m , η_1 — соответственно коэффициенты загрузки ЦУО при совмещенном (m — число каналов совмещения) и последовательном контроле ($m = 1$). Поэтому построенная модель АСК в виде одноканальной замкнутой системы массового обслуживания должна быть использована для оценки значения v в зависимости от числа совмещаемых программ m , их временных характеристик, определяемых параметрами λ и μ при фиксированной дисциплине диспетчеризации: «первым пришел — первым обслужен».

Коэффициент загрузки ЦУО при последовательной организации контроля η_1 определяется следующим образом:

$$\eta_1 = \frac{t_{ЦУО}}{t_{ЦУО} + t_{ФУ}} = \frac{\rho}{\rho + 1}; \quad \rho = t_{ЦУО} / t_{ФУ} = \lambda / \mu.$$

Введение мультипрограммного режима, с одной стороны, приводит к увеличению коэффициента загрузки η_m , но, с другой стороны, — к увеличению простоев в выполнении отдельных программ, что в общем нежелательно, причем степень увеличения загрузки и простоев программы зависит от числа совмещаемых программ. Согласно выбранной модели одноканальной замкнутой СМО коэффициент загрузки η_m определяется как

$$\eta_m = 1 - P_0 = 1 - \left[\sum_{j=0}^m \frac{m!}{(m-j)!} \rho^j \right]^{-1},$$

а средний коэффициент простоя программ

$$\varepsilon_m = \sum_{j=2}^m \frac{j-1}{m} \frac{m!}{(m-j)!} \rho^j P_0 = P_0 \sum_{j=2}^m \frac{(j-1)(m-1)!}{(m-j)!} \rho^j.$$

Тогда показатель эффективности введения мультипрограммной организации контроля v определяется следующим выражением:

$$v = \left\{ 1 - \left[\sum_{j=0}^m \frac{m!}{(m-j)!} \rho^j \right]^{-1} \right\} \cdot \frac{\rho + 1}{\rho}.$$

Зависимости η_m , ε_m и v (рис. 4-10) позволяют просто оценить эффективность мультипрограммной организации контроля от числа совмещаемых программ m и параметров программ ρ .

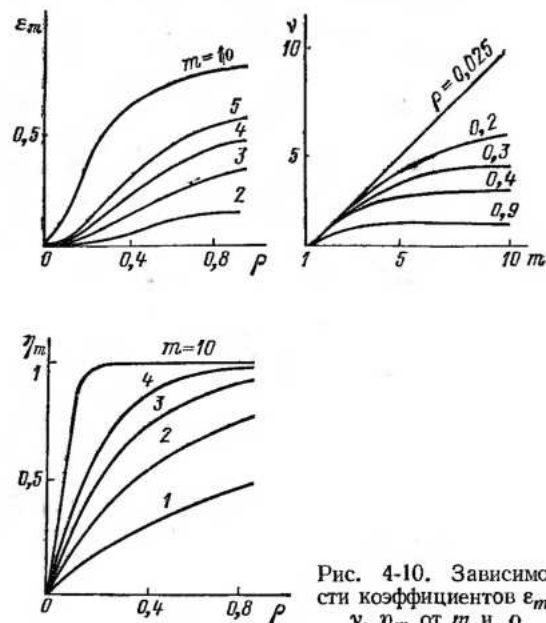


Рис. 4-10. Зависимости коэффициентов ε_m , v , η_m от m и ρ

Как показал предыдущий анализ, допущения о простейшем потоке требований, о показательнораспределенной длительности обслуживания и о наличии беспriorитетных дисциплин обслуживания: «первым пришел — первым обслужен» — приводят к тому, что функционирование многих систем, представимых в виде моделей СМО, описывается в классе марковских случайных процессов, в том числе процессов типа «гибель и размножение». В этом случае составление уравнений и получение из них выводов и расчетных формул оказываются достаточно простыми, как это было видно из рассмотренных примеров. Попытка отказаться хотя бы от одного из высказанных предположений

приводит к появлению значительных математических трудностей и требует применения специальных математических методов, разработанных в теории массового обслуживания, причем получение простых аналитических формул, доступных для инженерного анализа, удается только в редких, простейших случаях.

Рассмотренные методы описания и анализа поведения БСУ на языке марковских процессов представляют собой удобный математический аппарат для исследования только в том случае, когда число возможных состояний системы сравнительно невелико. В случае, когда число возможных состояний системы велико, эти методы перестают быть удобными по следующим причинам. Совместное решение большого числа не только дифференциальных, но и алгебраических уравнений оказывается затруднительным даже при использовании вычислительных машин. Кроме того, если даже нам удастся решить эти уравнения и найти вероятности всех состояний системы, полученные результаты могут быть трудно обзримыми.

Весьма перспективными с точки зрения анализа динамики БСУ при большом числе элементов оказываются методы, основанные на представлении структуры БСУ в виде сети приборов массового обслуживания и последующей декомпозиции задачи анализа общих характеристик сложной сети на задачи исследования характеристик простейших систем массового обслуживания, являющихся узлами сети [30], а также методы динамики средних, позволяющие получать средние характеристики системы, минуя определение вероятности состояния системы [8]. Краткое изложение данных подходов является содержанием следующей главы.

ГЛАВА ПЯТАЯ

МЕТОДЫ АНАЛИЗА ПОВЕДЕНИЯ СИСТЕМ ПРИ БОЛЬШОМ ЧИСЛЕ ЭЛЕМЕНТОВ

5-1. ПРЕДСТАВЛЕНИЕ СИСТЕМ В ВИДЕ СТОХАСТИЧЕСКИХ СЕТЕЙ. РАЗОМКНУТЫЕ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ

При проектировании БСУ ее структурную модель, отображающую взаимосвязи между отдельными операциями и звеньями управления в общем процессе функционирования

системы, удобно представлять в виде сети. Сеть является средством формализации процессов управления, она дает возможность: четко отобразить структуру исследуемого процесса; выявить с любой степенью детализации последовательность работ, образующих единый комплекс; отобразить в наглядной форме взаимосвязи и логическую последовательность выполнения работ; оценить степень важности каждого звена; исследовать продолжительность выполнения комплекса работ в целом; проанализировать возможные варианты достижения заданного результата и выбрать из них оптимальный по определенной системе показателей эффективности. Так как процессы, протекающие в БСУ, обычно носят случайный характер, то в дальнейшем будут рассматриваться только стохастические сети. При этом звенья системы обычно формализуются в виде подсистем массового обслуживания, а взаимосвязь между ними в общем процессе функционирования отображается в виде направленного графа. Вершины графа — подсистемы массового обслуживания, дуги графа — пути передачи требований информационных потоков между подсистемами массового обслуживания, составляющими сеть.

Для задания любой сети массового обслуживания необходимо определить: 1) источник требований, поступающих в сеть, имеющий некоторую интенсивность λ_0 ; 2) список обслуживаемых подсистем S_i ($i = 1, 2, \dots, M$), входящих в сеть; 3) вектор состава подсистем $m = (m_1, m_2, \dots, m_i, \dots, m_M)$, где m_i — число приборов, входящих в i -ю подсистему; 4) вектор интенсивностей обслуживания $\mu = (\mu_1, \mu_2, \dots, \mu_i, \dots, \mu_M)$, где μ_i — интенсивность обслуживания требования одним прибором, входящим в состав i -й подсистемы; 5) матрицу передач $R = \|r_{ij}\|$ ($i = 1, 2, \dots, M + 1$; $j = 1, 2, \dots, M + 1$), отражающую структуру передач информационных потоков.

При построении матрицы передач источник рассматривается как нулевая подсистема S_0 . Элемент матрицы r_{ij} характеризует вероятность выбора направления для требования в j -ю подсистему при окончании обслуживания в i -й подсистеме. При этом имеет место следующее соотношение:

$$\sum_{i=0}^M r_{ij} = 1, \quad i = 0, 1, 2, \dots, M.$$

Элемент $r_{00} = 0$, так как требования не остаются в источнике, а всегда поступают в систему.

Различают два вида сетей массового обслуживания: разомкнутые и замкнутые. В разомкнутых сетях интенсивность источника требований $\lambda_0 \neq 0$, т. е. имеет место постоянное взаимодействие с внешней средой. В замкнутых сетях массового обслуживания $\lambda_0 = 0$, интенсивность внешнего источника равна 0; следовательно, размерность матрицы передач R равна $M \times M$ (где M — число узлов сети). В замкнутых сетях задается всегда начальное конечное число требований N , которое все время сохраняется постоянным в данной сети, и в процессе функционирования сети изменяется только распределение требований по подсистемам, что и определяет динамику системы.

Аналитическое исследование возможно для стохастических сетей со следующими допущениями: 1) сеть формируют подсистемы с показательным законом (экспоненциальным) длительности обслуживания для каждого из приборов, входящего в подсистему; 2) сеть нагружена пуассоновским потоком однородных требований; 3) в каждой подсистеме массового обслуживания, входящей в сеть, очередь требований неограниченна, дисциплина обслуживания: «первым пришел — первым обслужен»; 4) сеть является линейной в том обычном смысле, что вероятность поступления требования в подсистему S_j за интервал времени $[t, t + \Delta t]$ является линейной комбинацией с постоянным коэффициентом r_{ij} вероятностей выхода требований из различных подсистем сети, включая и подсистему с номером j . Для линейных экспоненциальных сетей будем рассматривать только установившийся режим как наиболее часто интересующий при решении практических задач. Для установившегося режима в разомкнутой сети интенсивность среднего суммарного потока на входе любой подсистемы равна средней суммарной интенсивности выходного потока из данной подсистемы, т. е. справедлива следующая система уравнений:

$$\lambda_j = \sum_{i=0}^M r_{ij} \lambda_i, \quad j=0, 1, 2, \dots, M, \quad (5-1)$$

где λ_i — суммарная интенсивность на выходе i -й подсистемы; λ_j — суммарная интенсивность на выходе j -й подсистемы; r_{ij} — вероятность поступления требований из i -й подсистемы по окончании обслуживания в j -ю подсистему.

В матричной форме можно записать данную систему уравнений следующим образом:

$$R^T \lambda = \lambda, \quad (5-2)$$

где R^T — транспонированная матрица передач, а λ — вектор-столбец интенсивностей потоков требований λ_i , проходящих через подсистемы в сети в установившемся режиме.

Уравнение (5-2) можно представить в несколько ином виде, используя понятие единичной матрицы E :

$$(R^T - E) \lambda = 0.$$

Решение данного матричного уравнения существует, если определитель $|R^T - E| = 0$. Так как $\sum_{j=0}^M r_{ij} = 1$ при $i = 0, 1, 2, \dots, M$, а также учитывая, что прибавление к элементам какой-либо строки элементов какой-либо другой строки, умноженных на одно и то же число, не изменяет значения определителя, можно легко показать, что данное условие выполняется.

Замыкание в разомкнутой сети происходит через источник требований, отображающий внешнюю среду; следовательно, можно считать, что каждый узел сети является возвратным, т. е. из любой подсистемы через некоторое число шагов можно попасть в любую другую подсистему, в том числе и в источник требований. Поэтому установившийся режим в такой сети существует (ранг матрицы $|R^T - E|$ равен M) и для каждого значения интенсивности источника требований λ_0 существует единственное решение системы уравнений (5-1):

$$\lambda_j = \alpha_j \lambda_0, \quad j = 1, 2, \dots, M, \quad (5-3)$$

где α_j — коэффициент передачи от источника требований к j -му узлу.

Условие отсутствия возможности образования бесконечной очереди в каждой из подсистем разомкнутой сети выглядит следующим образом:

$$\alpha_j \lambda_0 < m_j \mu_j, \quad j = 1, 2, \dots, M,$$

что и определяет ограничение на интенсивность требований, поступающих из источника:

$$\lambda_0 < \min_i (m_i \mu_i / \alpha_i). \quad (5-4)$$

Для иллюстрации применения системы уравнения (5-1) рассмотрим пример.

Пример 5-1. Пусть задана разомкнутая стохастическая сеть (рис. 5-1).

Матрица передач R выглядит для данной сети следующим образом

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 \\ r_{10} & 0 & r_{12} & r_{13} \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

а интенсивность источника $\lambda_0 = 5 \text{ с}^{-1}$. Тогда в соответствии с (5-1) имеем следующую систему уравнений:

$$\left. \begin{aligned} \lambda_0 &= r_{10}\lambda_1; \\ \lambda_1 &= \lambda_2 + \lambda_3 + \lambda_0; \\ \lambda_2 &= \lambda_1 r_{12}; \\ \lambda_3 &= \lambda_1 r_{13}, \end{aligned} \right\}$$

откуда следует, что $\alpha_1 = 1/r_{10}$, $\alpha_2 = r_{12}/r_{10}$, $\alpha_3 = r_{13}/r_{10}$.

Рассмотрим далее вопросы определения основных характеристик как разомкнутых сетей в целом, так и их узлов в установившемся режиме.

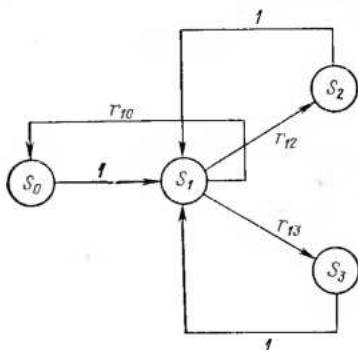


Рис. 5-1. Представление вычислительной системы в виде сети

При определении характеристик сети используется понятие «состояние сети». Под состоянием сети понимается вектор $n = (n_1, n_2, \dots, n_M)$, в соответствии с которым в 1-й подсистеме содержится n_1 требований, во 2-й подсистеме n_2 требований, в M -й подсистеме — n_M требований. Поэтому наиболее полной характеристикой сети является вероятность $P(n_1, n_2, \dots, n_M)$. Наиболее важным аналитическим инструментом для анализа характеристик сетей является теорема разбивания Джексона [30], содержащая достаточные условия, при выполнении которых разомкнутая сеть общего вида является совокупностью независимых полнодоступных систем с очередью. Ниже приводится теорема Джексона без доказательства, которое приведено в [30].

Теорема Джексона. Если для рассматриваемой сети выполнены условия:

1) все входящие потоки пуассоновские; 2) все маршруты переходов требований между подсистемами управляются матрицей вероятностей передач; 3) все длительности обслуживания имеют экспоненциальное распределение; 4) дисциплина выбора заявок из очереди в каждой подсистеме не зависит от длительностей обслуживания и маршрута заявок, то

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = \prod_{i=1}^M P_i(n_i);$$

$$P_i(n_i) = P_i(0) \prod_{i=1}^{n_i} \frac{\lambda_i}{\mu_i(i)};$$

$$\lambda_i = \alpha_i \lambda_0, \mu_i(i) = \begin{cases} i\mu_i, & i \leq m_i; \\ m_i\mu_i, & i > m_i. \end{cases}$$

Интерпретация вышеуказанных соотношений для разомкнутой сети, когда все узлы сети являются одноканальными системами массового обслуживания, выглядит следующим образом:

$$P_j(n_j) = P_j(0) \rho_j^{n_j}, \quad (5-5)$$

где $\rho_j = \alpha_j \lambda_0 / \mu_j$, $P_j(0) = 1 - \rho_j$;
а для многоканальных узлов:

$$P_j(n_j) = \begin{cases} P_j(0) \frac{\lambda_j^{n_j}}{n_j! \mu_j^{n_j}}, & n_j \leq m_j; \\ P_j(0) \frac{\lambda_j^{n_j}}{m_j! \mu_j^{n_j} m_j^{n_j - m_j}}, & n_j > m_j; \end{cases} \quad (5-6)$$

$$P_j(0) = \left[\sum_{n_j=0}^{m_j-1} \rho_j^{n_j} \frac{1}{n_j!} + \frac{\rho_j^{m_j}}{m_j! (1 - \rho_j/m_j)} \right]^{-1};$$

$$\rho_j = \lambda_j / \mu_j = \alpha_j \lambda_0 / \mu_j.$$

В соответствии с рассмотренным принципом декомпозиции задачи анализа сети в целом на задачи анализа отдельных подсистем имеем:

1) среднее число требований, находящихся в сети,

$$\bar{n} = \sum_{j=1}^M \bar{n}_j;$$

2) среднее число требований, ожидающих обслуживания в сети,

$$\bar{l} = \sum_{j=1}^M \bar{l}_j;$$

3) среднее время пребывания требования в сети

$$\bar{t} = \sum_{j=1}^M \alpha_j \bar{t}_j;$$

4) среднее время ожидания требования в сети

$$\bar{t}_{\text{ож}} = \sum_{j=1}^M \alpha_j \bar{t}_{\text{ож}j},$$

где $\alpha_j = \lambda_j / \lambda_0$ определяет среднее число прохождений требования через подсистему S_j , соответствующее одному прохождению через источник S_0 . Формулы для определения $\bar{n}_j$, $\bar{l}_j$, $\bar{t}_j$, $\bar{t}_{\text{ож}j}$ и других параметров, характеризующих отдельные подсистемы, приведены в § 4-3.

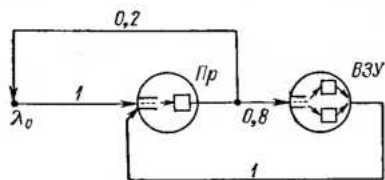


Рис. 5-2. Представление вычислительной подсистемы в виде разомкнутой сети

бразжает работу вычислительной подсистемы в составе процессора (Пр) и двух внешних однородных запоминающих устройств (ВЗУ), управляемых в составе мультиплексного канала. Пусть матрица передач выглядит следующим образом:

$$R = \begin{bmatrix} 0 & 1 & 0 \\ 0,2 & 0 & 0,8 \\ 0 & 1 & 0 \end{bmatrix};$$

$$\lambda_0 = 0,1 \text{ с}^{-1}; \mu_1 = 2 \text{ с}^{-1}; \mu_2 = 1 \text{ с}^{-1}.$$

Система уравнений (5-1) выглядит следующим образом:

$$\left. \begin{aligned} \lambda_0 &= 0,2\lambda_1; \\ \lambda_1 &= \lambda_0 + \lambda_2; \\ \lambda_2 &= 0,8\lambda_1, \end{aligned} \right\}$$

откуда $\lambda_1 = 5 \lambda_0 = 0,5 \text{ с}^{-1}$ и $\lambda_2 = 0,8 \cdot 5 \lambda_0 = 0,4 \text{ с}^{-1}$.

В рассматриваемой сети стационарный режим существует, так как условие (5-4) выполняется:

$$0,1 < \min \left(\frac{2}{5}; \frac{2 \cdot 1}{4} \right).$$

Среднее время пребывания требования в системе S_1 определяется как

$$\bar{t}_1 = \frac{1}{\mu_1} + \frac{\lambda_1}{\mu_1(\mu_1 - \lambda_1)} = \frac{1}{\mu_1 - \lambda_1} = \frac{1}{2 - 0,5} = \frac{1}{1,5} = 0,67 \text{ с}.$$

Среднее время пребывания требования в системе S_2 определяется как

$$\bar{t}_2 = \frac{1}{\mu_2} + \frac{\rho^2 P_0}{2\mu_2 2! \left(1 - \frac{\rho}{2}\right)^2};$$

$$P_0 = \left[1 + \rho + \rho^2 \frac{1}{2!} + \frac{\rho^3}{2! (2 - \rho)} \right]^{-1};$$

$$\bar{t}_2 = 1 + \frac{(0,4)^2}{2 \cdot 1 \cdot 2! \cdot \left(1 - \frac{0,4}{2}\right)^2} \cdot 0,67 = 1,04 \text{ с}.$$

Среднее время пребывания требования в сети

$$\bar{t} = \sum_{j=1}^2 \alpha_j \bar{t}_j = 5 \cdot 0,67 + 4 \cdot 1,04 = 7,51 \text{ с}.$$

Анализ характеристик систем, структура которых отображается в виде многофазных СМО, можно рассматривать как частный случай разомкнутых сетей массового обслуживания.

5-2. ЗАМКНУТЫЕ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ И АНАЛИЗ ИХ ХАРАКТЕРИСТИК

В замкнутых сетях задается всегда конечное число требований, которое сохраняется постоянным в данной сети, а в процессе функционирования изменяется только распределение требований по подсистемам.

Аппарат замкнутых сетей массового обслуживания получил в настоящее время широкое применение для анализа временных характеристик вычислительных комплексов как многомашинных, так и многопроцессорных в больших автоматизированных системах сбора и обработки информации.

Для замкнутой сети должны обязательно выполняться условия: $r_{i0} = 0$, $r_{0j} = 0$, $i = 0, 1, \dots, M$, $j = 0, 1, 2, \dots, M$, так как требования не уходят из сети, а просто перерас-

пределяются по подсистемам. В этом случае матрица передач имеет размерность $M \times M$ и система уравнений, эквивалентная системе уравнений (5-1), выглядит следующим образом:

$$\lambda_j = \sum_{i=1}^M r_{ij} \lambda_i, \quad j = 1, 2, \dots, M. \quad (5-7)$$

Следовательно, для существования решения определитель матрицы $|R^T - E|$ должен быть равен нулю. Можно показать также, что, как и для разомкнутых систем, в этом случае ранг матрицы $R^T - E$ равен $M - 1$, т. е. в установившемся режиме справедлива следующая система решений:

$$\lambda_j = \alpha_j \lambda_1, \quad \alpha_j = \lambda_j / \lambda_1$$

— при условии, что интересующие нас интенсивности λ_j выражены через интенсивность λ_1 . В общем случае решение системы уравнений (5-7) может быть выражено через интенсивность λ_i любой i -й подсистемы сети. Так как в замкнутой сети содержится конечное число требований, то в любой подсистеме может образоваться только конечная очередь. Поэтому анализ данных установившегося режима позволяет выявить только наиболее нагруженную подсистему. Наиболее нагруженной подсистемой в замкнутой сети считается подсистема, имеющая

$$\max_i \left\{ \frac{\alpha_j}{m_j \mu_j} \right\},$$

где $\alpha_j = \lambda_j / \lambda_1$, $j = 2, 3, \dots, M$, $\alpha_1 = 1$.

Для определения вероятностей состояний замкнутой сети используется тот же подход, что и для определения вероятностей разомкнутой сети, строгое доказательство которого приведено в [15], т. е. вероятности состояний сети выражаются через вероятности состояний составляющих сеть систем массового обслуживания. Однако полная декомпозиция задачи анализа характеристик замкнутых сетей невозможна, так как в сети содержится конечное число требований. Для учета этого факта вводится дополнительный множитель A , учитывающий, что для любых

состояний сети $\sum_{j=1}^M n_j = N$. В соответствии с этим

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = A \prod_{j=1}^M P_j(n_j).$$

Так как

$$\sum_{\sum_{j=1}^M n_j = N} P(n_1, n_2, \dots, n_j, \dots, n_M) = 1,$$

$$A = \left[\sum_{\sum_{j=1}^M n_j = N} \prod_{j=1}^M P_j(n_j) \right]^{-1};$$

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = \frac{\prod_{j=1}^M P_j(n_j)}{\sum_{\sum_{j=1}^M n_j = N} \prod_{j=1}^M P_j(n_j)}. \quad (5-8)$$

Из формул (5-5) и (5-6) следует, что вероятности $P_j(n_j)$ как для одноканальных, так и многоканальных систем могут быть представлены: $P_j(n_j) = d_j(n_j) P_j(0)$, причем для одноканальных систем

$$d_j(n_j) = \rho_j^{n_j}, \quad n_j = 1, 2, \dots,$$

а для многоканальных систем

$$d_j(n_j) = \begin{cases} \rho_j^{n_j} \frac{1}{n_j!}, & n_j \leq m_j; \\ \rho_j^{n_j} \frac{1}{m_j! m_j^{n_j - m_j}}, & n_j > m_j. \end{cases}$$

С учетом этого из формулы (5-8) имеем

$$\begin{aligned} P(n_1, n_2, \dots, n_j, \dots, n_M) &= \frac{\prod_{j=1}^M d_j(n_j) \prod_{j=1}^M P_j(0)}{\prod_{j=1}^M P_j(0) \sum_{\sum_{j=1}^M n_j = N} \prod_{j=1}^M d_j(n_j)} = \\ &= \frac{\prod_{j=1}^M d_j(n_j)}{\sum_{\sum_{j=1}^M n_j = N} \prod_{j=1}^M d_j(n_j)}. \end{aligned}$$

Можно показать, что вероятность $P(n_1, n_2, \dots, n_j, \dots, n_M)$ не зависит также от значения интенсивности λ_1 при условии, что все другие интересующие нас интенсивности λ_j выражены через интенсивность λ_1 (в общем случае через λ_i):

$$P(n_1, n_2, \dots, n_j, \dots, n_M) = \frac{\lambda_1^N \prod_{j=1}^M C_j(n_j)}{\lambda_1^N \sum_{\sum_{j=1}^M n_j = N} \prod_{j=1}^M C_j(n_j)} = \frac{\prod_{j=1}^M C_j(n_j)}{\sum_{\sum_{j=1}^M n_j = N} \prod_{j=1}^M C_j(n_j)}, \quad (5-9)$$

причем для одноканальных систем

$$C_j(n_j) = (\alpha_j / \mu_j)^{n_j}, \quad (5-10)$$

а для многоканальных систем

$$C_j(n_j) = \begin{cases} (\alpha_j / \mu_j)^{n_j} \frac{1}{n_j!}, & n_j \leq m_j; \\ (\alpha_j / \mu_j)^{n_j} \frac{1}{m_j! m_j^{(n_j - m_j)}}, & n_j > m_j. \end{cases} \quad (5-11)$$

Для $j = 1$ $\alpha_j = 1$.

В выражениях (5-9) — (5-11) удобно вместо параметра α_j ввести другой параметр ω_j , характеризующий вероятность поступления некоторого помеченного требования при его очередном переходе в j -ю подсистему и определяемый как

$$\omega_j = \frac{\alpha_j}{1 + \sum_{k=2}^M \alpha_k}.$$

Для этого необходимо числитель и знаменатель выражения (5-9) умножить на $\left[1 + \sum_{k=2}^M \alpha_k\right]^N$ и провести простейшие преобразования. Можно показать, что значения ω_j могут быть также определены непосредственно из системы

уравнений, подобной (5-7):

$$\omega_j = \sum_{i=1}^M r_{ij} \omega_i, \quad j = 1, 2, \dots, M$$

с учетом нормирующего условия

$$\sum_{j=1}^M \omega_j = 1.$$

Использование (5-9) для определения вероятностей $P(n_1, n_2, \dots, n_M)$ оказывается неудобным с вычислительной точки зрения при больших значениях N, M в силу трудностей, связанных с раскрытием суммы знаменателя. Поэтому используются вычислительные алгоритмы, легко реализуемые на ЭВМ. Рассмотрим два таких алгоритма применительно к замкнутым сетям с одно- и многоканальными системами массового обслуживания.

Одноканальные замкнутые сети. Введем следующее обозначение:

$$\sum_{\sum_{j=1}^r n_j = k} \prod_{j=1}^r C_j(n_j) = G_r(k), \quad (5-12)$$

тогда

$$\sum_{\sum_{j=1}^M n_j = N} \prod_{j=1}^M C_j(n_j) = G_M(N). \quad (5-13)$$

Для одноканальных СМО в узлах сети $C_j(n_j) = (\omega_j / \mu_j)^{n_j}$.

Обозначим $\omega_j / \mu_j = y_j$.

Тогда

$$\begin{aligned} G_r(k) &= \sum_{\substack{\sum_{j=1}^r n_j = k, \\ n_r = 0}} \prod_{j=1}^r y_j^{n_j} + \sum_{\substack{\sum_{j=1}^r n_j = k, \\ n_r > 0}} \prod_{j=1}^r y_j^{n_j} = \\ &= \sum_{\sum_{j=1}^{r-1} n_j = k} \prod_{j=1}^{r-1} y_j^{n_j} + y_r \sum_{\sum_{j=1}^{r-1} n_j = k-1} \prod_{j=1}^{r-1} y_j^{n_j} = \\ &= G_{r-1}(k) + y_r G_{r-1}(k-1), \end{aligned} \quad (5-14)$$

причем

$$\left. \begin{aligned} G_1(k) &= y_1^k, & k &= 0, 1, 2, \dots, N; \\ G_r(0) &= 1, & r &= 1, 2, \dots, M. \end{aligned} \right\} \quad (5-15)$$

Рекуррентное правило (5-14) совместно с начальными условиями (5-15) задает вычислительный алгоритм определения знаменателя в выражении (5-9) для вероятностей состояний замкнутых сетей с одноканальными СМО. Для иллюстрации данного принципа рассмотрим пример.

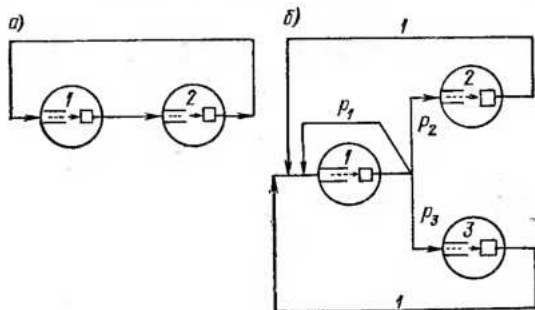


Рис. 5-3. Двухузловая сеть (а) и представление вычислительной системы в виде замкнутой сети (б)

Пример 5-3. Определить коэффициент загрузки процессора в вычислительной системе (рис. 5-3, б), работающей в мультипрограммном режиме.

Структура сети отображает взаимодействие процессора (1) с двумя устройствами внешней памяти (2; 3). Все подсистемы являются одноканальными.

Матрица передач R равна

$$R = \begin{bmatrix} P_1 & P_2 & P_3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

а число программ $N = 3$, число узлов $M = 3$.

Из решения системы уравнений

$$\left. \begin{aligned} \omega_2 &= \omega_1 P_2; \\ \omega_3 &= \omega_1 P_3; \\ \omega_1 &= \omega_1 P_1 + \omega_2 + \omega_3 \end{aligned} \right\}$$

с учетом нормировочного условия $\sum_{i=1}^3 \omega_i = 1$ получаем:

$$\omega_1 = \frac{1}{P_2 + P_3 + 1}; \quad \omega_2 = \frac{P_2}{P_2 + P_3 + 1}; \quad \omega_3 = \frac{P_3}{P_2 + P_3 + 1},$$

а значения $y_j = \omega_j / \mu_j$, μ_j считаются заданными, причем $y_i^N = C_i(n_j)$.

Коэффициент загрузки процессора в вычислительной системе определяется следующим образом:

$$\begin{aligned} \eta &= 1 - P(n_1=0, n_2=n_3) = \\ &= 1 - \sum_{j=1}^2 \sum_{n_j=3} \frac{C_2(n_2) C_3(n_3)}{\sum_{i=1}^3 C_1(n_1) C_2(n_2) C_3(n_3)} = \\ &= 1 - \frac{\sum_{j=1}^2 \sum_{n_j=3} C_2(n_2) C_3(n_3)}{\sum_{i=1}^3 \sum_{n_i=3} C_1(n_1) C_2(n_2) C_3(n_3)} = 1 - \frac{G_2^{(2,3)}(3)}{G_3(3)}. \end{aligned} \quad (5-16)$$

Рассмотрим многошаговый процесс вычисления величин $G_3(3)$. В соответствии с (5-14), (5-15)

1-й шаг:

$$G_2(1) = G_1(1) + y_2 G_2(0) = y_1 + y_2;$$

2-й шаг:

$$G_2(2) = G_1(2) + y_2 G_2(1) = y_1^2 + y_2(y_1 + y_2);$$

3-й шаг:

$$G_2(3) = G_1(3) + y_2 G_2(2) = y_1^3 + y_2 y_1^2 + y_2^2 y_1 + y_2^3;$$

4-й шаг:

$$G_3(1) = G_2(1) + y_3 G_3(0) = y_1 + y_2 + y_3;$$

5-й шаг:

$$G_3(2) = G_2(2) + y_3 G_3(1) = y_1^2 + y_2(y_1 + y_2) + y_3(y_1 + y_2 + y_3);$$

6-й шаг:

$$G_3(3) = G_2(3) + y_3 G_3(2) = y_1^3 + y_2 y_1^2 + y_2^2 y_1 + y_2^3 + y_3 y_1^2 + y_3 y_2^2 + y_3^2 y_1 + y_3^3 + y_3 y_1 + y_3 y_2 + y_3^2.$$

Аналогичным образом может быть подсчитано значение $G_3^{(2,3)}(3)$ при условии, что все требования перераспределяются по 2-му и 3-му узлам:

$$G_3^{(2,3)}(3) = y_2^3 + y_3 y_2^2 + y_3^2 y_2 + y_3^3.$$

По определенным значениям $G_3(3)$, $G_3^{(2,3)}(3)$ может быть подсчитан коэффициент загрузки η в соответствии с формулой (5-16). Аналогично может быть построен подобный вычислительный алгоритм и для замкнутых многоканальных сетей.

Многоканальные замкнутые сети. Рекуррентный алгоритм вычисления $G_M(N)$ в этом случае выглядит следующим

щим образом:

$$\begin{aligned} G_r(k) &= \sum_{\sum_{j=1}^r n_j = k} \prod_{j=1}^r C_j(n_j) = \\ &= \sum_{l=0}^k C_r(l) \sum_{\sum_{j=1}^{r-1} n_j = k-l} \prod_{j=1}^{r-1} C_j(n_j) = \\ &= \sum_{l=0}^k C_r(l) G_{r-1}(k-l) \end{aligned} \quad (5-17)$$

при начальных условиях:

$$G_1(k) = C_1(k) = \begin{cases} (\omega_1/\mu_1)^k \frac{1}{k!}, & k \leq m_1; \\ (\omega_1/\mu_1)^k \frac{1}{m_1! m_1^{k-m_1}}, & k > m_1; \end{cases}$$

$$G_r(0) = 1, \quad r = 1, 2, \dots, M.$$

Рассмотренный рекуррентный подход может быть обобщен не только на определение вероятностей $P(n_1, n_2, \dots, n_M)$, но и на другие характеристики сети как в целом, так и для отдельных узлов сети.

Введем следующие обозначения для средних значений показателей качества обслуживания в j -м узле сети при условии пребывания в сети N требований: $\bar{t}_{N,j}$ — среднее время пребывания требования; $\bar{t}_{N,j}^{\text{ож}}$ — среднее время ожидания требования обслуживания; $\bar{u}_{N,j}$ — среднее число занятых приборов; $\bar{n}_{N,j}$ — среднее число требований в узле; $\bar{l}_{N,j}$ — среднее число требований в очереди на обслуживание.

Из [30] известно, что

$$\bar{t}_{N,j} = \bar{t}_{N,j}^{\text{ож}} + 1/\mu_j; \quad (5-18)$$

$$\bar{n}_{N,j} = \bar{l}_{N,j} + \bar{u}_{N,j}. \quad (5-19)$$

При наличии стационарного режима в сети справедливо следующее тождество [15]:

$$\bar{t}_{N,j}/\bar{n}_{N,j} = \bar{t}_{N,j}^{\text{ож}}/\bar{l}_{N,j}. \quad (5-20)$$

Определяя из (5-20) $\bar{l}_{N,j} = \bar{n}_{N,j} \bar{t}_{N,j}^{\text{ож}}/\bar{t}_{N,j}$ и подставляя в (5-19), получим:

$$\bar{n}_{N,j} = \bar{u}_{N,j} + \bar{n}_{N,j} \bar{t}_{N,j}^{\text{ож}}/\bar{t}_{N,j},$$

откуда

$$\bar{t}_{N,j}^{\text{ож}} = \frac{\bar{n}_{N,j} \bar{t}_{N,j} - \bar{u}_{N,j} \bar{t}_{N,j}}{\bar{n}_{N,j}}. \quad (5-21)$$

Подставляя (5-21) в (5-18), получим:

$$\bar{t}_{N,j} = \frac{\bar{n}_{N,j} \bar{t}_{N,j} - \bar{u}_{N,j} \bar{t}_{N,j}}{\bar{n}_{N,j}} + \frac{1}{\mu_j},$$

откуда

$$\bar{t}_{N,j} = \frac{\bar{n}_{N,j}}{\bar{u}_{N,j}} \frac{1}{\mu_j}. \quad (5-22)$$

Выражение (5-22) в дальнейшем будет использоваться при определении основных показателей как сетей в целом, так и их узлов.

Определение среднего числа занятых приборов в узле. Для j -го узла сети оно будет

$$\begin{aligned} \bar{u}_{N,j} &= \sum_{n_j=1}^{m_j} n_j P_{N,j}(n_j) + m_j \sum_{n_j=m_j+1}^N P_{N,j}(n_j) = \\ &= \sum_{n_j=1}^{m_j} n_j \sum_{N, n_j} P(n_1, n_2, \dots, n_j, \dots, n_M) + \\ &+ m_j \sum_{n_j=m_j+1}^N \sum_{N, n_j} P(n_1, n_2, \dots, n_j, \dots, n_M), \end{aligned} \quad (5-23)$$

где $P_{N,j}(n_j)$ — вероятность нахождения в j -м узле сети n_j требований при условии, что во всей сети содержится N требований; $\sum_{N, n_j} P(n_1, n_2, \dots, n_j, \dots, n_M)$ — сумма вероятностей

по всем состояниям сети при условии, что в j -м узле находится n_j требований.

Подставляя в (5-23) выражение (5-9) и учитывая (5-13), получим:

$$\begin{aligned} \bar{u}_{N,j} &= [G_M(N)]^{-1} \cdot \left[\sum_{n_j=1}^{m_j} n_j \sum_{N, n_j} \prod_{i=1}^M C_i(n_i) + \right. \\ &\left. + m_j \sum_{n_j=m_j+1}^N \sum_{N, n_j} \prod_{i=1}^M C_i(n_i) \right]. \end{aligned} \quad (5-24)$$

Раскрывая $C_i(n_i)$ в соответствии с (5-11) и подставляя вместо α_j/μ_j значение ω_j/μ_j , после простейших преобразований получим:

$$\bar{u}_{N,j} = \frac{\omega_j}{\mu_j} [G_M(N)]^{-1} \cdot \left\{ \sum_{n_j=0}^{m_j-1} \sum_{N-1, n_j}^M \prod_{i=1}^M C_i(n_i) + \right. \\ \left. + \sum_{N-1, n_j=m_j}^M \prod_{i=1}^M C_i(n_i) + \sum_{n_j=m_j+1}^{N-1} \sum_{N-1, n_j}^M \prod_{i=1}^M C_i(n_i) \right\}.$$

В соответствии с (5-13) выражение в фигурных скобках есть не что иное, как $G_M(N-1)$, и, следовательно,

$$\bar{u}_{N,j} = \frac{\omega_j}{\mu_j} \frac{G_M(N-1)}{G_M(N)}. \quad (5-25)$$

Для одноканальных систем выражение (5-25) позволяет определить коэффициент загрузки обслуживающего прибора в j -м узле. Для иллюстрации применения данного выражения рассмотрим пример.

Пример 5-4. Задана одноканальная сеть с двумя узлами (рис. 5-3, а), число требований в сети $N=1$. Определить коэффициент загрузки в каждом из узлов.

В соответствии с (5-25)

$$\eta_j = \frac{\omega_j G_2(0)}{\mu_j G_2(1)}.$$

Так как $\omega_1 = \omega_2 = 1/2$, а в соответствии с (5-14), (5-15)

$$G_2(1) = G_1(1) + y_2 G_2(0) = y_1 + y_2 = \frac{\omega_1}{\mu_1} + \frac{\omega_2}{\mu_2},$$

то

$$\eta_j = \frac{1/\mu_j}{1/\mu_1 + 1/\mu_2},$$

что соответствует физическим представлениям о загрузке приборов в подобной простейшей сети.

Определение среднего времени пребывания требования в j -м узле. Согласно формуле (5-22) имеем

$$\mu_j \bar{t}_{N,j} = \bar{n}_{N,j} / \bar{u}_{N,j}.$$

Отсюда в соответствии с (5-9) и (5-13) получим

$$\mu_j \bar{t}_{N,j} = \frac{[G_M(N)]^{-1}}{\bar{u}_{N,j}} \cdot \left[\sum_{n_j=1}^N n_j \sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right].$$

Раскрывая $\bar{u}_{N,j}$ в соответствии с формулой (5-25), после простейших преобразований получим:

$$\mu_j \bar{t}_{N,j} = \frac{1}{\frac{\omega_j}{\mu_j} G_M(N-1)} \left\{ \sum_{n_j=1}^{m_j} n_j \left[\sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right] + \right. \\ \left. + \sum_{n_j=m_j+1}^N n_j \left[\sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right] + \right. \\ \left. + m_j \sum_{n_j=m_j+1}^N \left[\sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right] - \right. \\ \left. - m_j \sum_{n_j=m_j+1}^N \left[\sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right] \right\} = \\ = \frac{1}{\frac{\omega_j}{\mu_j} G_M(N-1)} \left\{ \sum_{n_j=1}^{m_j} n_j \left[\sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right] + \right. \\ \left. + m_j \sum_{n_j=m_j+1}^N \left[\sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right] + \right. \\ \left. + \sum_{n_j=m_j+1}^N (n_j - m_j) \left[\sum_{N, n_j}^M \prod_{i=1}^M C_i(n_i) \right] \right\} = \\ = \frac{1}{G_M(N-1)} \left\{ \sum_{n_j=0}^{N-1} \left[\sum_{N-1, n_j}^M \prod_{i=1}^M C_i(n_i) \right] + \right. \\ \left. + \frac{1}{m_j} \left[\sum_{n_j=m_j}^{N-1} (n_j - m_j + 1) \sum_{N-1, n_j}^M \prod_{i=1}^M C_i(n_i) \right] \right\} = \\ = \frac{1}{G_M(N-1)} \left[G_M(N-1) + \right. \\ \left. + \frac{1}{m_j} \sum_{n_j=m_j}^{N-1} (n_j - m_j + 1) P_{N-1,j}(n_j) G_M(N-1) \right] = \\ = 1 + \frac{1}{m_j} \left[\sum_{n_j=m_j}^{N-1} (n_j - m_j + 1) P_{N-1,j}(n_j) \right]; \\ \bar{t}_{N,j} = \frac{1}{\mu_j} \left[1 + \frac{1}{m_j} \sum_{n_j=m_j}^{N-1} (n_j - m_j + 1) P_{N-1,j}(n_j) \right]. \quad (5-26)$$

Для определения вероятностей $P_{N-1,j}(n_j)$ можно получить простые рекуррентные соотношения. Рассмотрим следующее отношение:

$$\begin{aligned} \frac{P_{N,j}(n_j)}{\bar{\mu}_{N,j}} &= \frac{[G_M(N)]^{-1}}{\bar{\mu}_{N,j}} \sum_{N,n_j} \prod_{i=1}^M C_i(n_i) = \\ &= \frac{1}{G_M(N-1)} \sum_{N,n_j} \prod_{i=1}^M C_i(n_i) = \\ &= \begin{cases} \frac{1}{n_j G_M(N-1)} \cdot \sum_{N-1, n_j-1} \prod_{i=1}^M C_i(n_i) = \frac{1}{n_j} P_{N-1,j}(n_j-1), \\ 1 \leq n_j \leq m_j; \\ \frac{1}{m_j G_M(N-1)} \cdot \sum_{N-1, n_j-1} \prod_{i=1}^M C_i(n_i) = \frac{1}{m_j} P_{N-1,j}(n_j-1), \\ n_j > m_j. \end{cases} \end{aligned}$$

Используя (5-25), получим:

$$P_{N,j}(n_j) = \frac{\omega_j G_M(N-1)}{\mu_j G_M(N)} \frac{P_{N-1,j}(n_j-1)}{\min(n_j, m_j)}. \quad (5-27)$$

Используя (5-27) и следующие начальные условия:

$$P_0(0) = 1; P_{N-k,j}(0) = \frac{G_M^{\bar{0}}(N-k)}{G_M(N-k)}; k = 1, 2, \dots, N-1 \quad (5-28)$$

(где значение $G_M^{\bar{0}}(N-k)$ вычисляется в предположении, что $N-k$ требований распределены по всем M узлам сети, за исключением узла j), можно вычислить все вероятности $P_{N-1,j}(n_j)$, входящие в выражение (5-27), а следовательно, и среднее время пребывания в j -м узле сети $t_{N,j}$, среднее время ожидания $t_{N,j}^{\text{ож}}$ в j -м узле сети. В соответствии с (5-20) и (5-22) определяются среднее число требований в узле $\bar{n}_{N,j}$ и среднее число требований в очереди на обслуживание $l_{N,j}$. Для одноканальных узлов замкнутой сети рассмотренные выше выражения значительно упрощаются.

Выражение (5-26) для $t_{N,j}$ при $m_j = 1$ ($j = 1, 2, \dots, M$) преобразуется следующим образом:

$$t_{N,j} = \frac{1}{\mu_j} \left[1 + \sum_{n_j=1}^{N-1} n_j P_{N-1,j}(n_j) \right] = \frac{1}{\mu_j} (1 + \bar{n}_{N-1,j}). \quad (5-29)$$

С другой стороны, среднее число требований $\bar{n}_{N,j}$ в j -м узле при наличии N требований в замкнутой сети

$$\bar{n}_{N,j} = N q_{N,j}, \quad (5-30)$$

где $q_{N,j}$ — вероятность пребывания требования в j -м узле сети, определяемая следующим образом:

$$q_{N,j} = \frac{\omega_j t_{N,j}}{\sum_{j=1}^M \omega_j t_{N,j}}. \quad (5-31)$$

Из (5-31) и (5-29) получим следующее рекуррентное выражение для определения $t_{N,j}$:

$$t_{N,j} = \frac{1}{\mu_j} \left[1 + \frac{(N-1) \omega_j t_{N-1,j}}{\sum_{j=1}^M \omega_j t_{N-1,j}} \right], \quad (5-32)$$

где $t_{0,j} = 0$.

Рассмотрим применение выражения (5-32) для анализа характеристики $t_{N,j}$ в примере 5-3 при условии, что $N = 2$.

Пример 5-4.

1-й шаг ($N = 1$):

$$t_{1,1} = 1/\mu_1; t_{1,2} = 1/\mu_2; t_{1,3} = 1/\mu_3;$$

2-й шаг ($N = 2$):

$$\begin{aligned} t_{2,1} &= \frac{1}{\mu_1} \left(\frac{\frac{1}{P_2+P_3+1} \cdot \frac{1}{\mu_1}}{\frac{1}{P_2+P_3+1} \cdot \frac{1}{\mu_1} + \frac{P_2}{P_2+P_3+1} \cdot \frac{1}{\mu_2} + \frac{P_3}{P_2+P_3+1} \cdot \frac{1}{\mu_3}} + 1 \right) = \\ &= \frac{1}{\mu_1} \left(\frac{\frac{1}{\mu_1}}{\frac{1}{\mu_1} + \frac{P_2}{\mu_2} + \frac{P_3}{\mu_3}} + 1 \right); \\ t_{2,2} &= \frac{1}{\mu_2} \left(\frac{\frac{1}{\mu_2} \cdot P_2}{\frac{1}{\mu_1} + \frac{P_2}{\mu_2} + \frac{P_3}{\mu_3}} + 1 \right); \quad t_{2,3} = \frac{1}{\mu_3} \left(\frac{\frac{1}{\mu_3} \cdot P_3}{\frac{1}{\mu_1} + \frac{P_2}{\mu_2} + \frac{P_3}{\mu_3}} + 1 \right). \end{aligned}$$

Если считать, что $P_1 = P_2 = P_3 = 1/3$, а $\mu_1 = \mu$, $\mu_2 = \mu/3$, $\mu_3 = \mu/3$, то $t_{2,1} = 4/(3\mu)$; $t_{2,2} = 4/\mu$; $t_{2,3} = 4/\mu$, что свидетельствует о том, что среднее время пребывания в узлах 2 и 3 в три раза больше времени пребывания в узле 1 (рис. 5-3, б) для рассматриваемой замкнутой одноканальной сети.

Используя выражения (5-30), (5-31), (5-32), (5-22), можно вычислить все интересующие нас основные характеристики в узлах замкнутой одноканальной стохастической сети:

1) среднее время пребывания в j -м узле:

$$t_{N,j} = \frac{1}{\mu_j} \left(1 + \frac{(N-1) \omega_j t_{N-1,j}}{\sum_{i=1}^M \omega_i t_{N-1,i}} \right);$$

$$t_{0,j} = 0, \quad j = 1, 2, \dots, M;$$

2) среднее число требований в j -м узле:

$$\bar{n}_{N,j} = N \omega_j t_{N,j} / \sum_{i=1}^M \omega_i t_{N,i};$$

3) среднее время ожидания в j -м узле:

$$t_{N,j}^{\text{ож}} = \frac{1}{\mu_j} \left(\frac{(N-1) \omega_j t_{N-1,j}}{\sum_{i=1}^M \omega_i t_{N-1,i}} \right);$$

4) среднее число требований в очереди в j -м узле:

$$l_{N,j} = \bar{n}_{N,j} t_{N,j}^{\text{ож}} / t_{N,j};$$

5) коэффициент загрузки обслуживающего прибора в j -м узле:

$$\eta_{N,j} = \frac{\bar{n}_{N,j}}{t_{N,j} \mu_j}.$$

Изложенные теоретические положения применения теории сетей к описанию и анализу БСУ являются только введением в аппарат теории сетей массового обслуживания. Все возрастающая сложность систем заставляет при использовании теории сетей развивать новые подходы к построению сетевых моделей. Одним из таких подходов является использование иерархии при построении сетевых моделей.

Большая информационная система в целом может быть представлена с помощью многоуровневой сетевой модели,

причем основная часть аппаратного оборудования и логические подсистемы на верхнем уровне отображаются укрупненной макроуровневой сетевой моделью, а параметры входящих в макроуровневую модель подсистем определяются из детализированных, или микроуровневых, сетевых моделей. Анализируя другие существующие подходы к описанию динамики БСУ, следует остановиться на одном из методов, который в настоящее время широко используется при построении моделей функционирования систем, — методе динамики средних.

5-3. МЕТОД ДИНАМИКИ СРЕДНИХ

Метод динамики средних позволяет осуществить изучение характеристик случайных процессов, протекающих в БСУ с большим числом состояний. При этом, в отличие от применения марковских процессов, он позволяет составить уравнения непосредственно для интересующих нас характеристик без рассмотрения вероятности состояний системы. Как отмечается в [8], основой применимости метода динамики средних является сложность изучаемых процессов и большое число участвующих в них элементов.

Пусть имеется система, состоящая из N элементов. Считаем, что элементы однородны и каждый из них может случайным образом переходить из состояния в состояние. Предположим, что все потоки событий, переводящие элемент из одного состояния в другое, — пуассоновские; тогда динамика поведения элемента будет описываться марковским случайным процессом. Обозначим число состояний элемента — n . В соответствии с ранее рассмотренным в § 4-1 состояние системы можно определить как совокупность состояний элементов и анализировать динамику системы, описывая ее марковским процессом. Однако в этом случае число состояний системы будет весьма большим при большом числе элементов N , поэтому даже составление системы уравнений будет затруднительным, не говоря уже о ее решении. В силу этого за основу метода динамики средних взят другой подход. Забудем на некоторое время о динамике системы и будем рассматривать только элемент. Введем понятие численности элементов j -го состояния, определяемое числом элементов в системе, находящихся в момент t в j -м состоянии, $X_j(t)$, $j = 1, 2, \dots, n$. Очевидно, что

$$\sum_{j=1}^n X_j(t) = N.$$

Изменение численности X_j во времени есть некоторый случайный процесс. Будем считать поведение случайного процесса определенным, если найдены математическое ожидание $M_j[t] = M[X_j(t)]$ и дисперсия $D_j[t] = D[X_j(t)]$. Для определения этих характеристик необходимо знать число элементов в системе и динамику элемента, представленную графом переходов. Введем переменную $X_j^{(i)}(t)$ для обозначения состояния i -го элемента: $X_j^{(i)} = 1$, если i -й элемент находится в j -м состоянии в момент t ; $X_j^{(i)} = 0$, если i -й элемент не находится в j -м состоянии в момент t .

Очевидно, что общая численность элементов, находящихся в j -м состоянии,

$$X_j(t) = \sum_{i=1}^N X_j^{(i)}(t).$$

Будем предполагать, что элементы независимы между собой; следовательно, каждый из них переходит из состояния в состояние независимо друг от друга и поэтому переменные $X_j^{(i)}$ ($i = 1, 2, \dots, N$) можно считать также независимыми между собой. Считаем, что динамика элементов одинакова, так как элементы однородны; тогда в соответствии с общей системой уравнений (4-5) для марковского процесса описания поведения элемента имеем

$$\left. \begin{aligned} P_j^{(i)'}(t) &= -P_j^{(i)}(t) \sum_{\substack{k=1, \\ j \neq k}}^n \lambda_{jk}^{(i)} + \sum_{\substack{k=1, \\ j \neq k}}^n P_k^{(i)}(t) \lambda_{kj}^{(i)}, \\ \sum_{j=1}^n P_j^{(i)}(t) &= 1; \quad P_1^{(i)}(0) = 1; \quad P_j^{(i)}(0) = 0, \\ &\quad j = 2, 3, \dots, n, \end{aligned} \right\} \quad (5-33)$$

где $P_j^{(i)}(t)$ — вероятность того, что i -й элемент в момент t находится в j -м состоянии.

Допуская, что значение $P_j^{(i)}$ уже определено из системы уравнений (5-33), имеем:

$$\begin{aligned} M_j^{(i)}[t] &= 0 \cdot [1 - P_j^{(i)}(t)] + 1 \cdot P_j^{(i)}(t) = P_j^{(i)}(t); \\ D_j^{(i)}[t] &= [0 - P_j^{(i)}(t)]^2 \cdot [1 - P_j^{(i)}(t)] + [1 - P_j^{(i)}(t)]^2 P_j^{(i)}(t) = \\ &= P_j^{(i)}(t) [1 - P_j^{(i)}(t)]. \end{aligned}$$

Так как $P_j^{(i)}(t) = P_j(t)$, $i = 1, 2, \dots, N$, то

$$M_j[t] = NP_j(t); \quad (5-34)$$

$$D_j[t] = NP_j(t) [1 - P_j(t)]. \quad (5-35)$$

Из выражений (5-34), (5-35) следует, что средние значения численностей элементов $M_j[t]$ и дисперсии численностей элементов $D_j[t]$ можно найти, не определяя вероятностей состояний системы в целом, а зная вероятность состояний отдельных элементов и число элементов в системе. Интересующие нас вероятности состояний элементов находятся в результате решения соответствующей системы дифференциальных уравнений (5-33), описывающей поведение одного элемента. Очевидно, что, умножая систему уравнений (5-33) на N и принимая во внимание (5-34), можно получить систему уравнений для определения средних численностей элементов $M_j[t]$:

$$\left. \begin{aligned} M_j' [t] &= - \left(\sum_{\substack{k=1, \\ j \neq k}}^n \lambda_{jk} \right) M_j[t] + \sum_{\substack{k=1, \\ j \neq k}}^n \lambda_{kj} M_k[t]; \\ M_1[0] &= N; \quad M_j[0] = 0, \quad j = 2, 3, \dots, n; \quad \sum_{j=1}^n M_j(t) = N. \end{aligned} \right\} \quad (5-36)$$

В этом случае дисперсия $D_j[t]$ определяется следующим образом:

$$D_j[t] = M_j[t] \left[1 - \frac{M_j[t]}{N} \right]. \quad (5-37)$$

Полученная система уравнений (5-36) называется системой уравнений динамики средних.

В ряде случаев интенсивности λ_{jk} , входящие в систему уравнений, не являются постоянными и зависят от того, какова численность элементов рассматриваемого состояния $X_j(t)$, а она является случайной величиной для рассматриваемого момента t . В некоторых ситуациях интенсивности λ_{jk} могут зависеть сразу от численности элементов нескольких состояний. Следовательно, в этом случае система уравнений (5-36) становится неопределенной и неразрешимой. Однако эту трудность обходят, если предположить, что интенсивности потоков событий, переводящих элемент из состояния в состояние, зависят не от самих численностей элементов $X_j(t)$, $j = 1, 2, \dots, n$, а от их средних значений $M_1[t]$, $M_2[t]$, ..., $M_n[t]$. Данное допущение называют принципом квазирегулярности, который позволяет получить приближенное решение составленных уравнений динамики средних (5-36). Указанное выше допущение приводит к значительным ошибкам при малых значениях N . Если же общее число элементов N велико, отклонение численности элементов каждого состояния от среднего значения относительно мало и метод динамики средних дает сравнительно малые погрешности.

Существенное значение имеет также вид зависимости, связывающей интенсивности потоков событий с численностями элементов состояний. Чем ближе эта зависимость к линейной, тем меньшую погрешность дает замена случайных численностей их средними значениями.

Поясним методику использования принципа квазирегулярности на простом примере, приведенном в [8]. Пусть система состоит из большого числа однородных технических устройств, каждое из которых может быть в одном из двух состояний: исправно (1) — работает, неисправно (2) — в ремонте. На каждый элемент (устройство) действует поток неисправностей с интенсивностью λ , не зависящей от численности элементов обоих состояний. Ремонт элементов занята группа рабочих в составе r человек ($r \leq N$). Каждый неисправный элемент ремонтируется одним рабочим, каждый рабочий может ремонтировать в среднем μ элементов в единицу времени. В начальный момент все элементы исправны. Все потоки событий — пуассоновские. Определим изменение средней численности устройств, находящихся в исправном состоянии $M_1[t]$, и соответственно средней численности устройств, находящихся в ремонте $M_2[t]$.

В данном случае для анализа динамики можно было воспользоваться рассмотренной уже моделью замкнутой многоканальной СМО, включающей в себя r каналов при N требованиях источника конечных требований. Однако для решения данной задачи используем метод динамики средних.

Сначала определим, какова будет средняя интенсивность потока ремонтов, приходящаяся на один ремонтируемый элемент:

$$\bar{\mu} = \varphi(X_2) = \begin{cases} \mu_{\Sigma}/X_2 = \mu X_2/X_2 = \mu & \text{при } X_2 \leq r; \\ \mu_{\Sigma}/X_2 = \mu r/X_2 & \text{при } X_2 > r, \end{cases}$$

где μ_{Σ} — суммарная интенсивность потока ремонтов, приходящаяся на все элементы, которые находятся в состоянии ремонта (численность

X_2). Построим граф динамики элемента (рис. 5-4); общая система дифференциальных уравнений для описания поведения элемента имеет вид:

$$\begin{cases} M_1' [t] = -\lambda M_1 [t] + \varphi (X_2) M_2 [t]; \\ M_2' [t] = -\varphi (X_2) M_2 [t] + \lambda M_1 [t], \end{cases}$$

где $M_1 (t)$, $M_2 (t)$ — средние численности элементов, находящихся в состояниях 1, 2.

Используя принцип квазирегулярности, можно перейти к следующей системе дифференциальных уравнений:

$$\begin{cases} M_1' [t] = -\lambda M_1 [t] + \varphi (M_2) M_2 [t]; \\ M_2' [t] = -\varphi (M_2) M_2 [t] + \lambda M_1 [t]. \end{cases}$$

Из этих двух уравнений мы можем выбрать одно, например второе, первое отбросить, а во второе подставить выражение $M_1 [t]$ из условия $M_1 [t] = N - M_2 [t]$; в результате получим

$$M_2' (t) = -\varphi (M_2) M_2 [t] + \lambda \{N - M_2 [t]\}, \quad M_2 [0] = 0.$$

Это — дифференциальное уравнение с разделяющимися переменными, которое легко может быть проанализировано с помощью известных методов интегрирования, в результате чего будет определена функция $M_2 [t]$, а следовательно, $M_1 [t]$.

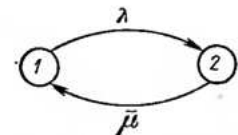


Рис. 5-4. Граф поведения элемента при использовании метода динамики средних

Метод динамики средних может использоваться не только для анализа БСУ с однородными элементами, но и с неоднородными элементами. Обобщение метода динамики средних для систем с неоднородными элементами, проведенное в [8], рассматривает два случая.

1. Система состоит из нескольких групп однородных элементов. В этом случае поведение каждой группы однородных элементов

описывается аналогичными системами дифференциальных уравнений типа (5-36), а их взаимосвязь, вытекающая из системы как единого целого, отражается в интенсивностях перехода. Интенсивности перехода в графе, отображающем поведение одной группы однородных элементов, являющихся функциями численностей элементов другой группы. Поведение групп элементов в системе, кроме того, определяется общими начальными условиями для всей системы.

2. Система состоит из некоторого «уникального элемента» и группы однородных элементов. Такой случай характерен для сложных систем, имеющих иерархическую структуру. В этом случае поведение группы однородных элементов описывается уравнениями типа (5-36) относительно средних численностей этих элементов, а поведение «уникального элемента» рассматривается с позиций марковской модели состояний, динамика которой отображается уравнениями вида (5-33); при этом взаимосвязь систем уравнений типа (5-36) и (5-33) осуществляется также через интенсивности перехода и начальные условия. В системе уравнений типа (5-36) интенсивности перехода зависят от состояния «уникального элемента», а в системе уравнений типа (5-33) — от численности элементов в определенных состояниях.

Естественно, что модели поведения систем в обоих случаях при наличии неоднородных элементов оказываются более сложными; системы

уравнений являются обычно нелинейными, и их решение требует применения численных методов, а следовательно, и использования ЭВМ. Кроме того, определение установившегося режима, как показано в [8], требует в каждом конкретном случае дополнительного исследования.

Метод динамики средних может быть использован и для описания поведения систем, в которых число элементов N в процессе функционирования системы не является постоянным, т. е. система не является замкнутой. В процессе функционирования системы число элементов, находящихся в определенных состояниях, может пополняться извне. Это пополнение очень легко учесть в уравнениях динамики средних. Для иллюстрации этого принципа рассмотрим пример.

Пример 5-5. Пусть задана система, состоящая из N_0 однородных элементов; граф состояний элемента показан на рис. 5-5. Если пополнения числа элементов в каком-то состоянии в ходе процесса не происходит, то уравнения динамики средних будут

$$\begin{cases} M_1' [t] = -(\lambda_{12} + \lambda_{13}) M_1 [t] + \lambda_{21} M_2 [t]; \\ M_2' [t] = -(\lambda_{21} + \lambda_{23} + \lambda_{24}) M_2 [t] + \lambda_{12} M_1 [t]; \\ M_3' [t] = +\lambda_{13} M_1 [t] + \lambda_{23} M_2 [t] + \lambda_{43} M_4 [t]; \\ M_4' [t] = -\lambda_{43} M_4 [t] + \lambda_{24} M_2 [t], \end{cases}$$

причем любое уравнение может быть отброшено и соответствующая переменная выражена из условия

$$\sum_{i=1}^4 M_i [t] = N_0.$$

Если теперь предположить, что число элементов, находящихся в одном из состояний (например, в 1), пополняется извне, причем интенсивность пополнения равна δ (среднее число элементов, вводимых в данное состояние за единицу времени), то первое уравнение рассматриваемой системы будет выглядеть следующим образом:

$$M_1' [t] = -(\lambda_{12} + \lambda_{13}) M_1 [t] + \lambda_{21} M_2 [t] + \delta,$$

а остальные уравнения сохраняются.

В общем случае величина δ может быть постоянной и переменной, как зависящей, так и не зависящей от средних численностей элементов каких-либо состояний. Начальное условие вида $\sum_{i=1}^4 M_i [t] = N_0$

также изменится. Ранее в любой момент времени сумма средних численностей элементов была равна одной и той же величине N_0 ; теперь она будет изменяющейся со временем численностью:

$$N(t) = N_0 + \int_0^t \delta(t) dt = N_0 + \delta t,$$

где N_0 — начальное значение численности элементов.

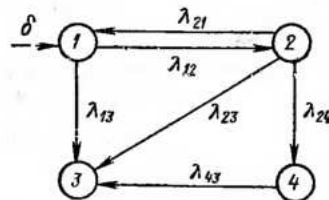


Рис. 5-5. Учет пополнения числа элементов при методе динамики средних

Таким образом, учет пополнения числа элементов сводится к тому, что к правой части соответствующего дифференциального уравнения прибавляется слагаемое, равное интенсивности пополнения — среднему числу элементов, вводимых в данное состояние элементов за единицу времени. Как видно из рассмотренного, метод динамики средних позволяет в удобной форме учитывать множество различных факторов, присущих динамике БСУ. Это определило его широкое применение при анализе эффективности, например, системы организации боевых действий, исследовании поведения различных экологических систем и характеристик больших информационно-вычислительных систем [26].

Методы, изложенные в данной главе, не исчерпывают всех возможных аналитических подходов, применяемых в настоящее время при описании поведения и анализе БСУ. По мнению авторов, они являются наиболее разработанными и доведенными до применения в широком классе задач анализа поведения БСУ различного назначения.

Содержание гл. 4 и 5 было ориентировано на изложение методов аналитического исследования поведения систем, в связи с чем были рассмотрены различные принципы построения моделей функционирования, допускающие такое исследование. Однако каждый раз при использовании того или иного подхода необходимо было делать целый ряд серьезных допущений, сужающих область его применения, но зато позволяющих построить более простую модель, а следовательно, быстрее получить результаты, интересующие системотехника на этапе макропроектирования всей системы.

В тех случаях, когда нас интересуют вопросы более полного описания функционирования системы и более детального анализа основных показателей качества ее функционирования, приходится отказываться от допущений и различного рода ограничений и строить более сложные модели, исследование которых возможно только путем имитационного моделирования поведения системы физическим и программным способами [4, 5, 33]. В настоящее время очень большое распространение получил метод программного имитационного моделирования, чему и будет посвящено содержание следующей главы.

ГЛАВА ШЕСТАЯ ПРИМЕНЕНИЕ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ ДЛЯ ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

6-1. ПОНЯТИЕ ОБ ИМИТАЦИОННОМ МОДЕЛИРОВАНИИ. ОСНОВНЫЕ ЕГО ЭТАПЫ

Имитационное моделирование является относительно новым и быстро развивающимся методом исследования поведения БСУ. Этот метод состоит в том, что с помощью ЭВМ воспроизводится поведение исследуемой БСУ, а исследователь-системотехник, управляя ходом процесса имита-

ции и обобщая получаемые результаты, делает вывод о ее свойствах и качестве поведения. Поэтому под имитацией следует понимать численный метод проведения на ЭВМ экспериментов с математическими моделями, описывающими поведение БСУ для определения интересующих нас функциональных характеристик. Появление имитационного моделирования и превращение его в эффективное средство анализа сложных систем было, с одной стороны, обусловлено потребностями практики, а с другой стороны, обеспечено развитием метода статистических испытаний (метода Монте-Карло), открывшего возможность моделирования случайных факторов, которыми изобилуют реальные системы, а также развитием электронной вычислительной техники, являющейся базой для проведения статистических экспериментов.

В развитие метода Монте-Карло, являющегося математической основой машинной имитации, значительный вклад был внесен работами Н. П. Бусленко, Д. И. Голенко, И. М. Соболя и др. [4, 13, 45]. Большое значение для развития машинной имитации также имела серия работ Н. П. Бусленко [4, 6], в которых на основе теоретического обобщения опыта статистического моделирования была разработана так называемая теория агрегатов, позволяющая строить имитационные модели широкого класса систем по общей схеме.

Достаточно широкое применение метода имитации при исследовании поведения БСУ обусловлено следующими причинами: а) сложностью модели поведения системы; наличием множества случайных факторов, которые ограничивают эффективность применения традиционных аналитических методов исследования, а в ряде случаев вообще исключают возможность их применения, в результате чего имитационное моделирование оказывается единственным способом исследования; б) новыми возможностями, которые позволяют осуществлять: наблюдение за поведением системы в таких условиях, в которых натурный эксперимент просто невозможен (либо в силу чисто физических причин, либо в силу ограниченности временных и стоимостных ресурсов); проведение имитационных экспериментов в широком диапазоне изменения параметров системы и внешней среды, что позволяет получить дополнительную полезную информацию в условиях информационной неопределенности, всегда сопутствующей начальным этапам решения системотехнических задач; прогнозирование поведения

системы позволяет получать к тому же ответ в сжатом масштабе времени; в) детальное наблюдение за поведением имитируемой системы позволяет лучше понять содержание самой системы и разработать такие предложения по ее улучшению, которые были бы невозможны без имитации; г) имитационное моделирование позволяет дать представление о том, какие из параметров системы являются наиболее существенными; д) имитационное моделирование может быть использовано как педагогический прием для обучения студентов и инженеров основным навыкам теоретического анализа, статистического анализа и теории принятия решений при управлении БСУ.

Но, как и у любого инструмента исследования, у метода имитации есть преимущества и недостатки. К недостаткам можно отнести: а) в ряде случаев имитационные модели оказываются достаточно сложными, что требует больших временных и стоимостных затрат на программирование, отладку моделей и проведение экспериментов; б) «имитационный мир», как и реальная действительность, оказывается трудно постижимым, ибо сложная имитационная модель приводит к такому числу разнообразных исходов, что в результате получаемую информацию не так легко интерпретировать; в) анализ результатов имитации основан только на использовании математической статистики, а, как известно, для получения статистической достоверности результатов, достаточной для обоснования выбора варианта управления, варианта построения системы и других, требуется многократное повторение имитационных экспериментов, что в ряде случаев требует больших временных затрат. Однако, несмотря на это, все равно остающиеся флуктуации результатов заставляют проявлять осторожность при подведении итогов машинных имитационных экспериментов; г) имитационное моделирование пока не располагает хорошо методически обоснованными принципами построения моделей для широкого класса БСУ, а поэтому каждый конкретный случай требует значительной специальной проработки.

Однако следует отметить, что, несмотря на отмеченные недостатки, метод имитационного моделирования как инструмент исследования БСУ вызывает большой научный интерес и в настоящее время интенсивно разрабатывается, поэтому можно надеяться, что многие из отмеченных недостатков этого метода будут устранены в ближайшее время.

При имитационном моделировании на ЭВМ можно выделить следующие основные этапы исследования: 1) формулировка проблемы; 2) построение математической модели функционирования системы; 3) составление и отладка программы для ЭВМ, включая и разработку процедур моделирования различных случайных факторов; 4) планирование имитационных экспериментов; 5) проведение экспериментов и обработка результатов исследования.

Рассмотрим более подробно содержание каждого из этапов.

1. *Формулировка проблемы* предполагает определение либо вопросов, на которые надо ответить, либо гипотез, которые надо проверить, либо воздействий, которые надо оценить, что в целом определяет цель имитации, в соответствии с которой должны быть определены и критерии, по которым оценивают результаты имитации.

2. *Построение математической модели* функционирования системы включает в себя определение входных, выходных, управляющих переменных и их взаимосвязи в общем алгоритме функционирования системы с целью оценки значений выбранных критериев. В случае машинной имитации математическая модель часто представляется в виде алгоритмического описания моделируемого процесса. Основой для этого является содержательное описание процесса функционирования системы. При построении модели необходимо учитывать два противоречивых фактора.

Усложнение модели, т. е. включение в модель большого числа переменных, приводит к большим временным затратам на составление, отладку модели, увеличивается и само время проведения имитационных экспериментов, а в некоторых случаях и возникают трудности с интерпретацией результатов. В результате может быть утрачена ценность полученных результатов в силу их большого времени запаздывания. Упрощение модели может привести к потере содержательности, модель становится неадекватной системе.

На этом этапе определяются, какие из переменных являются случайными, какие детерминированными, и после определения структуры модели производится оценка значений ее параметров, чему предшествует этап сбора необходимой исходной информации. Данный этап должен обязательно закончиться проверкой адекватности модели объекту. Общей методики проверки адекватности не существует. Модель считается адекватной объекту исследования при наличии утвердительных ответов на ряд вопросов:

а) нет ли в модели несущественных переменных, которые не улучшают способность предсказывать поведение системы? б) все ли существенные входные и управляющие переменные включены в модель? в) правильно ли сформулированы функциональные связи между входными и выходными переменными системы? г) верно ли определены параметры модели? д) являются ли оценки случайных параметров построенной модели статистически значимыми?

Правильное и обоснованное решение задач 1 и 2-го этапов во многом определяет успех имитационного эксперимента.

3. Составление машинной программы предполагает решение следующих задач: а) составление самой программы с использованием как универсальных алгоритмических языков, так и проблемно-ориентированных на решение задач имитации; б) разработка программных процедур имитации различных случайных факторов, имеющих в системе; в) отладка программы.

4. При планировании экспериментов решаются следующие основные задачи: а) выбор способов ускорения сходимости статистических оценок интересующих нас критериев к истинным значениям; б) определение объема имитационных экспериментов; в) составление плана проведения машинных экспериментов, что особенно важно при решении задач оптимизации на основе имитации. Решение вышеуказанных задач и составляет содержание этапа планирования экспериментов.

5. Проведение экспериментов и обработка результатов в основном преследуют цель: используя все многообразие статистических критериев и максимум информации, полученной в процессе эксперимента, сделать выводы по результатам имитационного эксперимента и определить их точность.

Перейдем к раскрытию содержания основных вопросов, связанных с разбором принципов построения математических моделей функционирования БСУ, ориентированных на использование ЭВМ, предполагая, что проблема исследования сформулирована и критерии оценки определены. Ранее было отмечено, что функционирование многих БСУ можно рассматривать как процесс перехода системы из состояния в состояние, причем изменению состояния всегда предшествует появление некоторого дискретного события, поэтому такие системы получили название «системы с дискретными событиями» [41].

6-2. ИМИТАЦИЯ ФУНКЦИОНИРОВАНИЯ СИСТЕМ С ДИСКРЕТНЫМИ СОБЫТИЯМИ

Класс БСУ, формализуемых в виде систем с дискретными событиями, достаточно широк и включает в себя большие информационно-управляющие системы, вычислительные системы, системы связи и др. При решении задачи программной имитации любой системы, в том числе и системы с дискретными событиями, составляется содержательное описание процесса функционирования, формализованное в виде математической модели. При этом определяются параметры модели, аппроксимирующие функции для детерминированных зависимостей, а также типовые законы распределения для случайных величин и аппроксимирующие таблицы частот экспериментальных данных.

При имитации функционирования систем на ЭВМ построенная математическая модель преобразуется в моделирующий алгоритм, в котором сохраняются логическая структура, последовательность протекания процесса во времени, характер и состав информации о состояниях процесса.

ЭВМ представляют собой устройства дискретного типа, а потому и моделирующий алгоритм должен являться дискретной аппроксимацией построенной математической модели функционирования системы. Особенность имитации поведения БСУ на ЭВМ сводится к определению правила развертывания квазипараллельных процессов функционирования множества элементов в системе в последовательный моделирующий алгоритм.

Прямой путь решения данной задачи весьма прост. Интервал времени $[0; T]$, в течение которого рассматривается работа системы, разбивается на интервалы длиной Δt , из-за чего данный способ решения получил название принципа Δt .

В пределах каждого интервала последовательно вычисляются приращения всех процессов в модели и производится, если это нужно, изменение состояния отдельных элементов модели. При достаточно малых Δt получаем хорошее приближение имитируемых процессов к процессам в реальной системе с параллельным выполнением операций. При таком способе построения моделирующего алгоритма точность моделирования достигается ценой больших затрат машинного времени. Обычно такой способ построения имитационных моделей используется при моделировании непрерывных динамических систем. Принцип Δt

является наиболее универсальным принципом построения моделирующих алгоритмов, хотя и наименее экономичным с точки зрения вычислений на ЭВМ.

Однако данный способ мало пригоден для решения задач имитации больших информационных систем, динамика которых состоит в переходе из состояния в состояние, причем в промежутках между переходами состояние системы остается неизменным. Каждый такой переход связан с наступлением некоторого события в системе, например приход входного либо управляющего дискретного сигнала, приход требования, отказ элемента, достижение некоторой характеристикой системы заданного порогового значения и др.

Анализ различных алгоритмов моделирования для такого класса систем, проведенный в [4], показал, что наиболее часто используется принцип особых состояний. При построении алгоритма имитации в соответствии с данным принципом функционирование системы, формализованное в математической модели, рассматривается как совокупность параллельно протекающих процессов, причем каждый процесс есть некоторая последовательность событий, с каждым событием связано изменение состояния системы. Событие, возникающее в системе, определяется как особое состояние. Процессы в общем случае не являются независимыми, а взаимодействуют между собой. Для иллюстрации данных понятий рассмотрим пример.

Пример 6-1. Пусть в систему массового обслуживания, состоящую из одного прибора для обслуживания, поступает поток требований. Если требование, придя в систему, застает прибор свободным, то оно занимает прибор и обслуживается в нем в течение некоторого времени, после чего покидает систему. В противном случае требование поступает в очередь к прибору, где будет дожидаться конца обслуживания всех ранее поступивших в систему требований.

В такой системе возможны два процесса: 1) процесс поступления требований; 2) процесс «захвата» прибора и обслуживания требований, причем процесс поступления требований инициирует второй процесс, если последний был приостановлен по причине отсутствия требований.

Под состоянием системы будем понимать, как и ранее, число требований, находящихся в системе. Из алгоритма функционирования видно, что состояние системы изменяется либо под влиянием событий процесса № 1 (приход требований), либо под влиянием событий процесса № 2 (захват и обслуживание). Анализ содержания данного примера показывает, что для развертывания совокупности параллельно протекающих процессов в последовательный необходимо упорядочить во времени моменты наступления событий каждого из процессов. Далее, сканируя по временной упорядоченной последовательности и имитируя в каж-

дый наступивший момент особого состояния все необходимые действия, заданные в содержательном описании процесса функционирования системы как реакция системы на событие, получим имитационную модель алгоритма функционирования системы для ЭВМ. В этом и состоит содержание принципа особых состояний.

С целью формализации принципа особых состояний определим для каждого выделенного процесса момент T_i наступления очередного события i -го процесса и, если таких процессов будет n , то выбор наиболее раннего момента наступления особого состояния определится в соответствии с операцией

$$T_r = \min_i T_i, \quad (6-1)$$

где r — номер процесса, в котором наступило ближайшее событие.

Моменты T_i называются моментами системного времени, в отличие от реального времени, в котором работает моделирующая ЭВМ.

Развертывание квазипараллельных процессов функционирования системы в последовательный в соответствии с принципом особых состояний называется диспетчеризацией по принципу узловых точек.

Рассмотрим структурную схему моделирующего алгоритма (рис. 6-1) и назначение основных операторов.

Оператор задания начальных условий A содержит: оператор задания начальных условий для моделируемого варианта A_1 , оператор задания начальных условий для одной реализации (одного имитационного эксперимента) A_2 .

Оператор определения очередного момента изменения состояния системы B находит момент наступления наиболее раннего события в соответствии с формулой (6-1) и определяет вид состояния системы, в которое она переходит в данный момент времени. Таким образом, оператор B содержит два массива: массив времени T_i и массив состояний Z_i ($i = 1, 2, \dots, n$).

Логический оператор C осуществляет переход по номеру наступившего события к соответствующему оператору D_i , имитирующему реакцию системы на событие. Оператор реакции D_i имеет следующие основные функции:

а) выполняет все необходимые операции, предусмотренные в алгоритме функционирования системы, как реакцию на данное событие; б) вычисляет и накапливает интересующую системотехника статистику по исследуемым характеристикам системы, если она относится к данному процессу;

в) определяет момент следующего наступления события в данном процессе и состояние системы и заносит их в соответствующие массивы для оператора B ; г) вычисляет и заносит в массивы оператора B моменты наступления событий и состояние в других процессах, если последние были приостановлены, а инициирование их возможно только событием данного процесса; в результате этого пассивные,

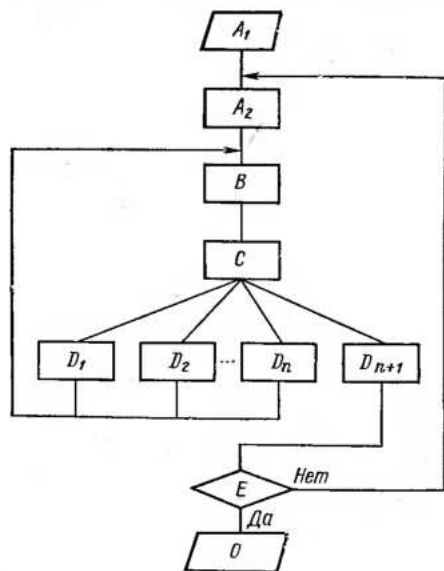


Рис. 6-1. Структурная схема моделирующего алгоритма поведения БСУ в соответствии с принципом особых состояний

заторможенные процессы становятся активными. Следует заметить, что если процесс переходит в приостановленное состояние, то время наступления очередного события определить нельзя. В этом случае обычно принимают $T_j = \infty$ (где j — номер такого процесса) для того, чтобы исключить этот процесс из массива T в силу соотношения (6-1). Данный процесс активизируется только в случае инициирования его событиями других процессов.

Оператор реакции D_{n+1} обычно вводится в тех случаях, когда имитация производится на ограниченном интервале длины T_0 , где $T_{n+1} = T_0$ — длительность одного имитационного эксперимента. Когда T_0 окажется очередным

моментом изменения состояния, управление передается блоку D_{n+1} , выполняющему все необходимые действия по завершению одного имитационного эксперимента.

Если не проведено достаточное для статистической точности число экспериментов, что проверяется оператором E , то осуществляется возврат к оператору A_2 ; в противном случае проводится обработка результатов в блоке оператора O (окончание) и выдача результатов имитационных экспериментов на устройство печати. Для конкретизации и более детального знакомства с данным подходом рассмотрим примеры.

Пример 6-2. Имитация поведения одноканальной СМО. В систему поступает однородный поток требований с заданным законом распределения интервалов. Длительность обслуживания — случайная величина с заданным законом распределения. На входе системы установлен накопитель. Если требование застает прибор занятым и накопитель заполненным, то оно теряется, получает отказ. В результате моделирования необходимо оценить вероятность отказа в обслуживании требований.

Структурная схема алгоритма моделирования приведена на рис. 6-2. Считаем, что число реализаций (имитационных экспериментов) нам задано; при этом под реализацией будем понимать процесс обработки одного требования. В схеме приняты следующие обозначения: $T_{ПОСТ}$ — момент времени поступления требования в систему; $ТОСВ$ — момент времени освобождения обслуживающего прибора; T — момент системного времени; TI — длительность интервала между моментами поступления соседних требований; $ТОБС$ — время обслуживания требования; K — количество требований в системе, находящихся на обслуживании и в очереди; $ОЧЕР$ — количество требований в очереди; $КОТК$ — количество требований, получивших отказ в обслуживании; $КОЛ$ — количество поступивших требований (текущее число реализаций); H — число заданных реализаций; $МОЧ$ — число мест для ожидания в очереди (в накопителе).

Оператор 1 осуществляет ввод заданных значений H , $МОЧ$ и присваивает нулевые значения счетчикам числа требований, поступивших в систему ($КОЛ$), находящихся в системе (K) и в очереди ($ОЧЕР$), получивших отказ в обслуживании ($КОТК$), а также переменным, имеющим значения текущего времени поступления требований ($T_{ПОСТ}$) и системного времени (T). Переменной, имеющей значение текущего времени освобождения прибора ($ТОСВ$), присваивается значение, равное машинной бесконечности (∞). Оператор 2 формирует случайный интервал между соседними поступающими требованиями в соответствии с заданным законом распределения. Оператор 3 формирует момент поступления первого требования в систему. Операторы 4, 5, 19 составляют ранее указанный оператор B , выбирающий очередной момент изменения состояния системы. Операторы 6—15 составляют обобщенный оператор реакции D_1 на приход требования в систему. Операторы 20—25 входят в обобщенный оператор реакции D_2 на освобождение обслуживающего прибора. Условный оператор 16 производит сравнение текущего числа требований ($КОЛ$), поступивших в систему с заданным числом (H). В случае, если условие не выполняется, управление передается оператору 4; в противном случае управление передается

оператору 17, осуществляющему обработку результатов. В данном случае оператор 17 вычисляет вероятность отказа в обслуживании, равную $KOTK/N$, и выводит на устройство печати результаты моделирования.

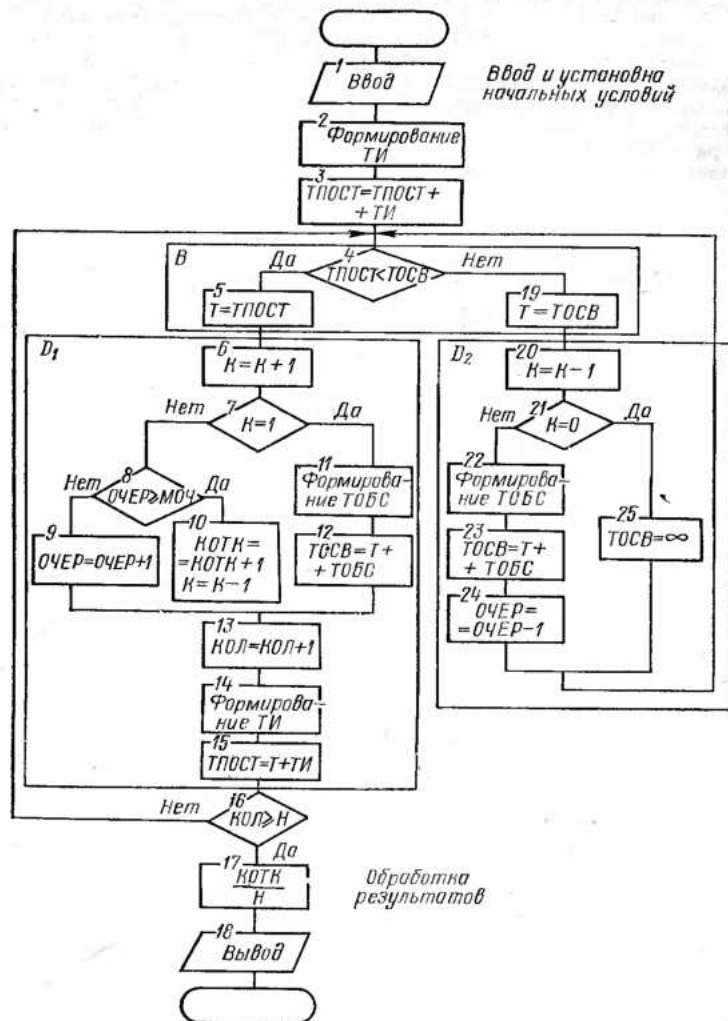


Рис. 6-2. Алгоритм моделирования одноканальной СМО

Хотя данный пример достаточно прост, он позволяет методически почувствовать основные особенности принципа имитации на основе выделения множества особых состояний.

Если при имитации динамики БСУ существуют однородные по характеру событий и реакций на них системы процессы, то их можно объединить в однородную группу и с каждой такой группой сопоставить один блок реакции; тогда общее число блоков реакции сократится. Однако при обслуживании одним блоком реакции группы однородных процессов следует предусмотреть настройку блока на соответствующий процесс. Иллюстрацию такого подхода рассмотрим на примере моделирования замкнутой сети массового обслуживания. Как было ранее показано, аналитическое исследование поведения сетей возможно только при определенных допущениях. Имитационное моделирование позволяет снять эти допущения и решать задачу анализа поведения БСУ, представленной сетью, в полном объеме.

Пример 6-3. Рассмотрим имитацию функционирования замкнутой сети массового обслуживания. Пусть задана замкнутая сеть массового обслуживания со следующими параметрами: 1) число требований в сети N ; 2) число подсистем массового обслуживания M , причем каждая подсистема является одноканальной; 3) матрица передач требований $R = \|r_{ij}\|$, на основе которой требование при окончании обслуживания в i -й подсистеме передается в j -ю подсистему с вероятностью r_{ij} ; 4) длительность обслуживания распределена по показательному закону с параметром μ_i ($i = 1, 2, \dots, M$); 5) дисциплина обслуживания в узлах сети: «первым пришел — первым обслужен». В результате имитации определить коэффициенты простоя обслуживающих подсистем в узлах сети.

При анализе функционирования данной сети можно выделить две группы однородных квазипараллельных процессов: 1) процессы окончания обслуживания требований в i -й подсистеме ($i = 1, 2, \dots, M$); 2) процессы поступления требований в j -ю подсистему ($j = 1, 2, \dots, M$) по окончании обслуживания в i -й подсистеме. При этом следует отметить, что моменты наступления событий первого процесса совпадают с моментами наступления событий второго процесса.

На структурной схеме алгоритма моделирования замкнутой сети массового обслуживания (рис. 6-3) введены следующие обозначения: $TA(I)$ ($I = 1, 2, \dots, 2M$) — элемент массива значений абсолютных моментов наступления событий в выделенной группе процессов, причем, если $I = 1, 2, \dots, M$, то элемент $TA(I)$ — момент времени освобождения I -й подсистемы; если $I = (M+1), (M+2), \dots, 2M$, то элемент $TA(I)$ — момент времени поступления требования в подсистему с номером $(I-M)$; T — момент системного времени; $K(I)$ ($I = 1, 2, \dots, M$) — элемент массива количества требований, находящихся в подсистемах массового обслуживания; $ТОБС(I)$ ($I = 1, 2, \dots, M$) — длительность обслуживания в I -й подсистеме сети, формируемая по показательному закону с параметром μ_i ($i = 1, 2, \dots, M$); $ТОСВ(I)$ ($I = 1, 2, \dots, M$) — элемент массива времен освобождения обслуживающих подсистем; $ТОЖ(I)$ ($I = 1, 2, \dots, M$) — элемент массива суммарного времени ожидания обслуживающих приборов в подсистемах сети, необходимый для определения коэффициентов простоя; $ТМОД$ — необходимое время имитации для накопления статистики; $ТП$ — последний момент системного времени, не превышающий значения $ТМОД$.

Рассмотрим работу алгоритма. Оператор 1 производит ввод основных исходных данных о параметрах сети и установку начальных значений. При этом предполагается, что все N требований находятся в первой подсистеме и отсчет времени моделирования начинается с момента завершения обслуживания первого требования в первой подсистеме, т. е. $TA(1) = 0$, $TA(I) = \infty$, $I = 2, 3, \dots, 2M$; $K(I) = N$, $K(I) = 0$,

$I \neq 1$. Всем элементам массива $ТОЖ(I)$ ($I = 1, 2, \dots, M$) присваиваются нулевые значения.

Операторы 2, 3 определяют очередной момент системного времени (T) и номер процесса (Π), в котором произошло событие, соответствующее данному моменту времени. Условный оператор 4 определяет момент времени окончания моделирования и выдачи результатов оператору

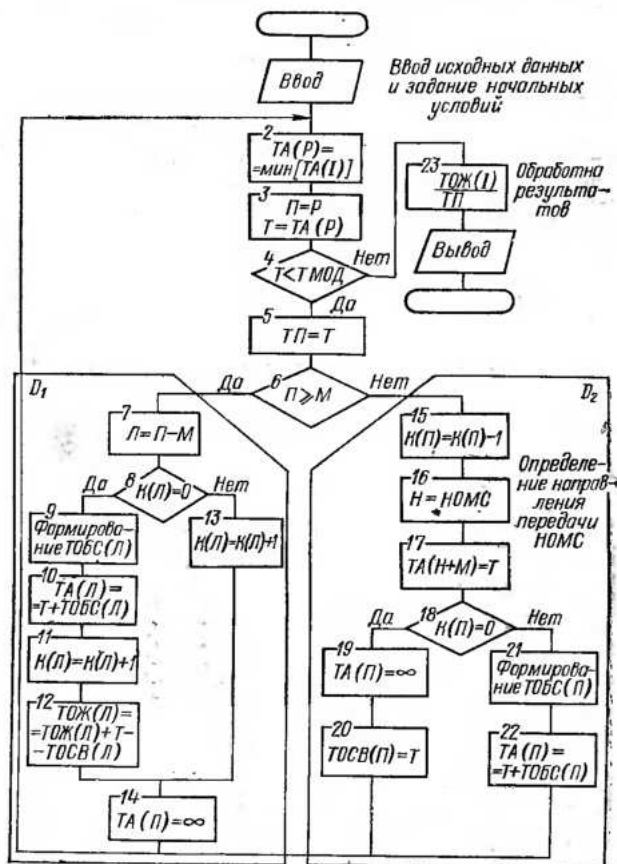


Рис. 6-3. Алгоритм моделирования замкнутой сети массового обслуживания

обработки 23, если условие оператора выполнено. Оператор 5 определяет предыдущий момент системного времени $TП$ для вычисления интегральных показателей качества процесса (коэффициентов простоев). Условный оператор 6 определяет вид события, наступившего в системе (освобождение прибора или приход требования), и переключает работу алгоритма на соответствующий оператор реакции. Обобщенный оператор реакции D_1 на приход требования реализует следующие алгоритмические действия:

а) определяет номер системы (Π), на которую пришло требование (оператор 7); б) проверяет: занят $[K(\Pi) \neq 0]$ либо свободен $[K(\Pi) = 0]$ прибор соответствующей обслуживаемой системы (оператор 8); в) в случае, если прибор свободен: формирует длительность обслуживания $ТОБС(\Pi)$ в соответствии с заданным законом распределения (оператор 9); определяет абсолютный момент освобождения прибора $ТА(\Pi)$ (оператор 10); фиксирует новое количество требований в системе $K(\Pi)$ (оператор 11); вычисляет и накапливает статистику по суммарным простоям обслуживающего прибора $ТОЖ(\Pi)$ (оператор 12); в случае, если прибор занят: ставит требование в очередь и тем самым формирует новое количество требований в системе (оператор 13); исключает данное требование из массива $ТА(I)$ ($I = 1, 2, \dots, 2M$), присваивается $ТА(\Pi) = \infty$ (оператор 14).

Обобщенный оператор реакции D_2 на событие, связанное с окончанием обслуживания, реализует следующие алгоритмические действия: а) фиксирует, что произошло обслуживание, т. е. изменяет количество требований в системе с номером Π (оператор 15); б) определяет направление передачи, т. е. номер системы $H = (НОМС)$, в которую поступает требование в соответствии с вероятностной матрицей передачи R при окончании обслуживания в системе с номером Π (оператор 16); в) фиксирует момент поступления требования в массиве $ТА(I)$ ($I = 1, 2, \dots, 2M$), присваивая переменной $ТА(H+M)$ значение, равное T (оператор 17); г) проверяет: освободился $[K(\Pi) = 0]$ или не освободился обслуживающий прибор $[K(\Pi) \neq 0]$ (оператор 18); в случае, если прибор освободился полностью, абсолютный момент освобождения прибора $ТА(\Pi)$ исключается из массива $ТА(I)$ ($I = 1, 2, \dots, 2M$) (оператор 19) и фиксируется момент освобождения прибора $ТОСВ(\Pi)$, равный T (оператор 20); в случае, если прибор не освободился полностью, формируется время обслуживания данного прибора $ТОБС(\Pi)$ в соответствии с заданным законом распределения (оператор 21) и определяется новый абсолютный момент освобождения прибора $ТА(\Pi)$ (оператор 22). После выполнения своих функций обобщенные операторы реакции D_1, D_2 передают управление оператору 2. Оператор 23 реализует функции обработки результатов моделирования, вычисляя коэффициенты простоя обслуживающих подсистем, равные $ТОЖ(I)/TП$ ($I = 1, 2, \dots, M$), где $ТОЖ(I)$ — суммарное время простоев i -й обслуживающей подсистемы; $TП$ — продолжительность времени имитации.

Рассмотренный принцип имитации поведения систем с дискретными событиями является основой построения имитационных моделей для решения различных задач, возникающих при проектировании БСУ; при этом структура и содержание обобщенных операторов реакции в каждом конкретном случае будут зависеть как от характера исследуемой системы, так и от цели проведения имитационных экспериментов. Общая же структура алгоритма моделирования останется той же. Данный принцип позволяет легко перейти к блочному представлению модели, когда можно включить в имитационный эксперимент как блоки, алгоритмическое содержание которых допускает аналити-

ческое исследование, так и блоки, допускающие экспериментальное исследование. Одновременно данный принцип является основой для обеспечения наглядности имитационной модели, если расчленение модели на операторы и блоки производится не только из вычислительных удобств, но из набора категорий, понятий, образов, привычных для специалистов, работающих в данной области исследования.

Так как при функционировании БСУ всегда имеется множество случайных факторов, то возникает задача их программной имитации на ЭВМ. В рассмотренных примерах к таким случайным факторам относятся: случайная длительность интервала между требованиями в потоке требований на обслуживание; случайная длительность обслуживания в системе; выбор направления передачи требования в соответствии с заданными вероятностными характеристиками. Поэтому далее рассмотрим основные алгоритмы программной имитации различных случайных факторов, сопровождающих процесс функционирования БСУ.

6.3. МЕТОДЫ ИМИТАЦИИ СЛУЧАЙНЫХ ФАКТОРОВ ПРИ ПРОГРАММНОМ МОДЕЛИРОВАНИИ СИСТЕМ

Базовой последовательностью случайных чисел, используемой для формирования в ЭВМ случайных элементов различной природы, с различными законами распределения, является совокупность случайных чисел с равномерным законом распределения

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & x < a, x > b. \end{cases}$$

где $f(x)$ — дифференциальный закон распределения равномерно распределенных чисел x в интервале $[a, b]$.

Строго говоря, на цифровой ЭВМ получить последовательность случайных величин с равномерным распределением не представляется возможным. Поэтому, если считать, что число разрядов ЭВМ равно k , а случайное число сформировано согласно формуле

$$\alpha = \sum_{j=1}^k \alpha_j 2^j,$$

где $\alpha_j = 0, P_j = 1/2$; $\alpha_j = 1, 1 - P_j = 1/2$, то $x = i/2^k - 1$ принимает значение $i/(2^k - 1)$, ($i = 0, 1, 2, \dots, 2^k - 1$) с вероятностью $P_i = 1/2^k$. Такое распределение чисел x называется квазиравномерным в интервале $[0, 1]$, причем математическое ожидание и дис-

персия определяются следующими соотношениями

$$\left. \begin{aligned} M[x] &= \sum_{i=0}^{2^k-1} \frac{i}{2^k-1} \cdot \frac{1}{2^k} = \frac{1}{2}; \\ D[x] &= \sum_{i=0}^{2^k-1} \left(\frac{i}{2^k-1} - \frac{1}{2} \right)^2 \cdot \frac{1}{2^k} = \frac{1}{12} \cdot \frac{2^k+1}{2^k-1}. \end{aligned} \right\} \quad (6-2)$$

Из формул (6-2) видно, что математическое ожидание $M[x]$ точно совпадает с генеральным средним для равномерного распределения в интервале $[0, 1]$, а дисперсия при $k \rightarrow \infty$ асимптотически стремится к дисперсии для равномерного распределения при $a = 0, b = 1$, равной $1/12$.

Практически при $k > 15$ обеспечивается требуемая точность в имитационных исследованиях. Поэтому в дальнейшем будем говорить о равномерном законе, хотя в действительности при программном моделировании имеем дело с квазиравномерным законом.

При выводе выражений (6-2) предполагалось, что x формируется на основе случайных чисел α_j , принимающих значения $(0; 1)$ с вероятностью $P_j = 1/2$, для чего в машине должен существовать случайный генератор, дающий строго случайные последовательности чисел α_j с соответствующим распределением. Так как в ЭВМ такого генератора нет, случайные числа вырабатываются программным путем, в силу чего они, строго говоря, не являются случайными, так как формируются на основе вполне детерминированных преобразований, поэтому их называют псевдослучайными. Такие последовательности случайных чисел являются периодическими, поэтому очень длинные последовательности, длина которых превосходит период, уже не будут строго случайными. Однако, если при моделировании число обращений к программному датчику случайных чисел оказывается меньше периода, измеряемого числом различных случайных чисел, то такая периодичность программного датчика не оказывает существенного влияния на результаты моделирования.

Основные достоинства программного способа получения псевдослучайных чисел состоят в следующем: а) не требуются специальные внешние устройства; б) получение чисел достаточно быстрое (обычно требуется 3—10 команд на число); в) возможно повторное воспроизведение чисел; г) требуется только однократная проверка алгоритма получения заданной последовательности чисел. Методы получения псевдослучайных квазиравномерных чисел программным путем можно разбить на две основные группы: а) аналитические; б) методы перемешивания. При использовании аналитических методов очередное число в псевдослучайной последовательности получается с помощью некоторого рекуррентного соотношения, аргументами которого являются одно или несколько предыдущих чисел последовательности

$$x_r = \varphi(x_{r-1}, x_{r-2}, \dots, x_0).$$

Простейшим примером указанного способа получения случайных чисел, равномерно распределенных в интервале $[0, 1]$, может служить метод вычетов [13], в котором используется следующее рекуррентное соотношение:

$$x_{i+1} = bx_i \pmod{M},$$

где выражение $bx_i \pmod{M}$ означает остаток от деления произведения bx_i на число M ; x_{i+1} — очередное случайное число; x_i — предыдущее случайное число; b — некоторая константа; M — число, определяющее наибольшее значение получаемых случайных чисел.

Данный способ является основой построения мультипликативного программного датчика случайных чисел. В этом случае алгоритм построения последовательности случайных чисел сводится к следующему: 1) выбрать в качестве параметра a_0 — произвольное нечетное число, например $a_0 = 5^{13}$ при разрядности ЭВМ $k = 32$, $M = 2^{31} - 1$; 2) вычислить коэффициент b по формуле $b = 8c \pm 3$, где c — любое целое положительное число, в частном случае $b = 5^{13}$; 3) вычислить произведение ba_0 ; взять k младших разрядов в качестве первого члена последовательности a_1 , остальные отбросить; 4) провести нормализацию числа по формуле $x_1 = a_1 / (2^k - 1)$; 5) вычислить очередное псевдослучайное число a_2 как k младших разрядов произведения ba_1 и вернуться к пункту 4.

Описанный генератор подвергался, как отмечено в [35], широкой экспериментальной проверке и проявил достаточно хорошие свойства. С другими аналитическими способами получения случайных чисел, равномерно распределенных в интервале $[0; 1]$, можно ознакомиться в [13].

В случае применения методов перемешивания очередное число последовательности получается путем хаотического перемешивания разрядов предыдущего случайного числа с помощью операций сдвига, специальных сложений и различных арифметических операций. Например, часто используются следующие комбинации операций для перемешивания разрядов предыдущего случайного числа: а) сдвиг предыдущего числа на некоторое число разрядов влево и специальное сложение результатов этого сдвига с предыдущим числом последовательности; б) сдвиг предыдущего числа на некоторое число разрядов влево и вправо и специальное сложение результатов этих сдвигов.

Данные операции заканчиваются взятием модуля и нормализацией. В качестве начальной константы для формирования последовательности обычно берут иррациональные числа ($\sqrt{3}/3$, $\sqrt{2}/2$, $\sqrt{5}/5$). Вообще в настоящее время не существует общей теории получения случайных чисел. Применение тех или иных методов получения случайных чисел во многом определяется тем типом ЭВМ, который используется для моделирования. Правомочность же применения того или иного способа получения случайных чисел программным путем определяется только результатом статистической проверки.

Для проверки качества серии квазиравномерных псевдослучайных чисел используются различные системы проверочных тестов. Укажем наиболее часто применяемые тесты.

Тест частот. Отрезок $[0; 1]$ разбивается на m (обычно 10—20) равных интервалов. Полученные эмпирические частоты n_i/N ($i = 1, 2, \dots, m$; $\sum_{i=1}^m n_i = N$) сравниваются с теоретическими вероятностями $1/m$. Согласно проверяется по критерию χ^2 , так как статистическая функция

$$\chi^2 = \frac{m}{N} \sum_{i=1}^m \left(n_i - \frac{N}{m} \right)^2$$

подчиняется распределению χ^2 с $(m - 1)$ степенями свободы, где N — объем выборки.

Тест пар. Рассматриваются последовательные пары случайных чисел при делении интервала $[0; 1]$ на m частей. Каждая пара случайно попадет в одно из m^2 делений квадратной таблицы $m \times m$. При этом в зависимости от метода образования пар изменяется число степеней свободы.

Пусть дана серия чисел $x_1, x_2, \dots, x_N$. Если пары образовывать в виде $(x_1, x_2), (x_3, x_4), \dots$, то пары взаимно независимы, эмпирические частоты (их число m^2) сравниваются с теоретическими вероятностями равномерного распределения $1/m^2$. Функция

$$\chi^2 = (2m^2/N) \sum_{i=1, j=1}^m (n_{ij} - N/2m^2)^2 \quad (6-3)$$

распределена по закону χ^2 с $(m^2 - m)$ степенями свободы, где n_{ij} — число попаданий в (i, j) -ю клетку таблицы размером $m \times m$; $N/2$ — объем выборки пар случайных чисел.

Более сложная ситуация возникает, если пары образовывать в виде $(x_1, x_2), (x_2, x_3), \dots, (x_i, x_{i+1})$. Этот метод образования пар более выгодный, так как полнее использует выборку чисел, но из-за зависимости пар распределение функции χ^2 уже другое. Как показано в [55], вместо (6-3) следует брать функцию

$$\chi^2 = \frac{m^2}{N} \sum_{i=1, j=1}^m \left(n_{ij} - \frac{N}{m^2} \right) - \frac{m}{N} \sum_{i=1}^m \left(n_i - \frac{N}{m} \right)^2,$$

которая, как и (6-3), распределена по закону χ^2 с $(m^2 - m)$ степенями свободы.

Тест частот направлен на проверку близости распределения полученной последовательности чисел к равномерной, а тест пар — на проверку независимости, а точнее, отсутствия корреляции в последовательности случайных чисел.

Тест на периодичность. Если среди множества программно-получаемых случайных чисел $x_1, x_2, \dots, x_{r-1}, \dots$ нет одинаковых, а следующее случайное число x_r совпадает с одним из полученных ранее чисел, то r называют отрезком аperiodичности. Очевидно, что $r \leq 2^k$, где k — число разрядов ЭВМ. При исследовании программного генератора случайных чисел всегда необходимо установить длину отрезка аperiodичности. Если число необходимых для имитационных экспериментов случайных чисел меньше отрезка аperiodичности r , то может быть использован датчик. В противном случае необходимо, строго говоря, использовать другой генератор случайных последовательностей. В настоящее время программное обеспечение любой современной ЭВМ содержит программы генераторов случайных последовательностей с проверенными статистическими свойствами и ориентированными на широкий класс имитационных экспериментов. Далее рассмотрим способы имитации различных случайных факторов на ЭВМ на базе последовательности псевдослучайных квазиравномерных чисел.

Имитация случайных событий. Пусть в результате эксперимента должно наступить одно из несовместных событий $A_1, A_2, \dots, A_k, \dots, A_n$, которые образуют полную группу событий, т. е. $\sum_{k=1}^n P_k = 1$, где P_k — вероятность наступления события A_k .

Разбиваем отрезок $[0; 1]$ на n частей длиной $P_1, P_2, \dots, P_n$; при этом точки деления отрезка имеют следующие координаты:

$$l_0 = 0; \quad l_1 = P_1; \quad l_2 = P_1 + P_2, \dots, l_n = \sum_{k=1}^n P_k = 1.$$

Пусть теперь x — очередное число от генератора случайных чисел. Если $l_{k-1} \leq x \leq l_k$, то считаем, что произошло событие A_k .

Действительно,

$$P(A_k) = P(l_{k-1} \leq x \leq l_k) = l_k - l_{k-1} = P_k.$$

Рассмотренная процедура может быть положена в основу выбора направления передачи требований при моделировании замкнутых сетей массового обслуживания, алгоритм которого был рассмотрен выше. Аналогичным образом можно моделировать дискретные случайные величины при конечном числе их значений. Если имеем дискретную случайную величину y , причем $y = 1$ с вероятностью P , а $y = 0$ с вероятностью $1 - P$, то при имитации ее на ЭВМ необходимо каждый раз решать следующую систему неравенств: если $0 \leq x_i \leq P$, то $y_i = 1$; если $P \leq x_i \leq 1$, то $y_i = 0$, где x_i — очередное случайное число от генератора случайных равномерно распределенных чисел.

Имитация непрерывных случайных величин. В литературе [41] рассматриваются несколько способов имитации, основанных на различных преобразованиях равномерно распределенных случайных чисел в числа с заданным законом распределения.

Метод обратной функции [41]. Он основан на использовании следующей теоремы.

Если x — случайная величина, равномерно распределенная на отрезке $[0; 1]$, то случайная величина y , являющаяся решением уравнения

$$\int_{-\infty}^y f(z) dz = x, \quad (6-4)$$

имеет плотность распределения $f(y)$.

Данный метод позволяет сформулировать правило генерирования случайных чисел, имеющих произвольное непрерывное распределение $f(y)$: 1) вырабатывается случайное число x_i генератором случайной равномерной последовательности; 2) случайное число y_i , имеющее распределение $f(y)$, находится из решения уравнения

$$\int_{-\infty}^{y_i} f(z) dz = x_i.$$

Таким образом, последовательность чисел $x_0, x_1, x_2, \dots, x_i$ преобразуется в последовательность чисел $y_0, y_1, y_2, \dots, y_i$, имеющую заданную плотность распределения $f(y)$. Рассмотрим примеры.

Пример 6-4. Необходимо получить последовательность чисел, равномерно распределенных на отрезке $[a, b]$. Тогда, используя (6-4), имеем:

$$x_i = \int_a^{y_i} \frac{1}{b-a} dz = \frac{y_i - a}{b-a},$$

откуда $y_i = x_i(b-a) + a$.

Пример 6-5. Необходимо получить последовательность чисел, имеющих распределение по показательной функции

$$f(y) = \begin{cases} \lambda e^{-\lambda y}, & 0 \leq y < \infty; \\ 0, & y < 0. \end{cases}$$

В соответствии с (6-4) имеем

$$x_i = \int_0^{y_i} \lambda e^{-\lambda z} dz = 1 - e^{-\lambda y_i},$$

откуда

$$y_i = -\frac{1}{\lambda} \ln(1 - x_i). \quad (6-5)$$

Так как величина $(1 - x_i)$ также имеет равномерное распределение на отрезке $[0; 1]$, то формула (6-5) может быть записана другим способом:

$$y_i = -\frac{1}{\lambda} \ln x_i.$$

Однако формула (6-4) не для всех распределений может быть использована по следующим причинам: а) зависимость $y_i = y(x_i)$ нельзя получить в явном виде; б) зависимость $y_i = y(x_i)$ является сложной для численных расчетов. В этом случае используют приближенные методы, например метод ступенчатой аппроксимации и предельные теоремы теории вероятности.

Метод ступенчатой аппроксимации. Зависимость плотности распределения $f(y)$ от возможных значений случайной величины y представляется графически в интервале изменения y от a до b . Если случайная величина задана на бесконечном интервале, то производим усечение распределения с заданной точностью. В данном случае указанная плотность $f(y)$ может быть получена также и экспериментально. Разобьем отрезок $[a, b]$ на n частей, таких, что

$$\int_{a_0}^{a_1} f(y) dy = \int_{a_1}^{a_2} f(y) dy = \dots = \int_{a_i}^{a_{i+1}} f(y) dy = \dots = \int_{a_{n-1}}^{a_n} f(y) dy = 1/n,$$

где a_i — координата точки разбиения ($i = 0, 1, 2, \dots, n$).

Тогда вероятность того, что случайная величина y попадет в один из интервалов,

$$P(a_i \leq y \leq a_{i+1}) = \int_{a_i}^{a_{i+1}} f(y) dy = \frac{1}{n} = \text{const},$$

т. е. попадание на любой отрезок $[a_i, a_{i+1}]$ случайной точки равновероятно. На каждом из интервалов функция $f(y)$ аппроксимируется ступенчатой функцией так, чтобы значение $f(y)$ в каждом интервале было постоянным; тогда координата случайной точки может быть представлена как $y_i = a_i + c_i$, где c_i — расстояние точки от левого конца интервала. В силу ступенчатой аппроксимации c_i является равномерно распределенной случайной величиной на интервале $[0; a_{i+1} - a_i]$. Правило имитации в этом случае сводится к следующему: а) получаем

два числа x_1, x_2 от генератора равномерно распределенных чисел; б) с помощью x_1 находим индекс $i = [nx_1]$ для интервала, где $[nx_1]$ — целая часть числа nx_1 , причем $[nx_1] \leq nx_1$; в) с помощью числа x_2 находим $c_i = x_2 (a_{i+1} - a_i)$; г) находим случайное число, имеющее интересующий нас закон распределения $f(y)$:

$$y = a_{[nx_1]} + x_2 (a_{[nx_1] + 1} - a_{[nx_1]}).$$

Таким образом, для получения случайного числа y , имеющего закон $f(y)$, используются два числа от генератора случайных чисел x_1, x_2 .

Использование предельных теорем. В некоторых случаях для имитации определенных законов распределения используют предельные теоремы теории вероятностей. Так, например, для получения нормального закона распределения используется свойство сходимости независимых величин к нормальному распределению. Метод обратной функции в этом случае оказывается неэффективным, так как получаемый при этом интеграл

$$x_i = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{y_i} \exp\left[-\frac{(z-a)^2}{2\sigma^2}\right] dz$$

не раскрывается в явную зависимость $y_i = y(x_i)$.

Для получения нормально распределенных чисел с параметрами $m_y = a = 0$, $\sigma_y = 1$ удобен искусственный прием, основанный на центральной предельной теореме теории вероятностей. Для этого в качестве исходных чисел возьмем n равномерно распределенных на отрезке $[-1; 1]$ чисел, получаемых из интервала $[0; 1]$ по правилу: $z_i = 2x_i - 1$.

Сформируем величину z согласно следующей формуле:

$$z = \sum_{i=1}^n z_i.$$

По центральной предельной теореме при достаточно большом значении n величина z может считаться нормально распределенной с параметрами

$$m_z = \sum_{i=1}^n M[z_i] = 0;$$

$$\sigma_z^2 = \sum_{i=1}^n \sigma_{z_i}^2 = n/3.$$

Проведя нормирование величины z , получим, что величина

$$u = \frac{z}{\sigma_z} = \frac{\sqrt{3}}{\sqrt{n}} \left[\sum_{i=1}^n (2x_i - 1) \right] \quad (6-6)$$

будет иметь нормальное распределение с параметрами $m_u = 0$ и $\sigma_u^2 = 1$. Практически установлено, что при $n \geq 8$ формула (6-6) дает вполне хорошие результаты.

Для ускорения процесса получения нормально распределенных случайных чисел (для уменьшения n) иногда вводят так называемую коррекцию. Нормально распределенные случайные числа с параметрами $m_y = 0$ и $\sigma_y^2 = 1$ могут быть, например, вычислены с помощью эмпирических формул

$$y_i = u_i - \frac{1}{20n} (3u_i - u_i^3); \quad (6-7)$$

$$y_i = u_i - \frac{41}{13440n^2} (u_i^5 - 10u_i^3 + 15u_i); \quad (6-8)$$

При использовании (6-7) достаточно принять $n = 5$, а при — (6-8) $n = 2$.

Для получения нормальной последовательности чисел s_i с параметрами $m_s = a$ и $\sigma_s^2 = \sigma^2$ достаточно выполнить линейное преобразование $s_i = \sigma u_i + a$. Другие приближенные способы получения случайных чисел с произвольными законами распределения подробно рассмотрены в [13].

Имитация дискретных случайных величин. Из всего множества законов распределения дискретных случайных величин рассмотрим наиболее часто встречающиеся в задачах имитации поведения БСУ:

а) величины y_i имеют биномиальное распределение

$$P_n(y_i = i) = C_n^i p^i (1-p)^{n-i}, \quad i = 0, 1, 2, \dots, n;$$

б) величины y_i имеют пуассоновское распределение с параметром a

$$P_n(y_i = i) = \frac{a^i}{i!} e^{-a}, \quad i = 0, 1, 2, \dots \quad (6-9)$$

В первом случае имитация величины y_i сводится к n -кратной имитации эксперимента с двумя исходами: $x_{ij} = 1$ с вероятностью P и $x_{ij} = 0$ с вероятностью $1 - P$ ($j = 1, 2, \dots, n$), что реализуется по уже рассмотренной выше схеме имитации дискретных случайных событий.

Тогда $y_i = \sum_{j=1}^n x_{ij}$ имеет распределение, близкое к биномиальному.

Во втором случае необходимо воспользоваться предельной теоремой Пуассона: если P_0 — вероятность наступления события A при одном испытании, то вероятность наступления i событий при n независимых испытаниях в случае, если $n \rightarrow \infty$ и $P_0 \rightarrow 0$, асимптотически стремятся к (6-9) при

$$\lim_{n \rightarrow \infty} (nP_0) = a$$

Поэтому имитация в этом случае проводится так же, как и в первом, только при условии, что

$$P \approx P_0 = a/n \leq 0,1. \quad (6-10)$$

Чем больше значение n , тем больше распределение чисел y_i будет приближаться к закону Пуассона (6-9). Значение n выбирается из условия (6-10) при известном параметре a .

Имитация потоков дискретных событий. Под потоком событий, как ранее было отмечено, понимают последовательность однородных событий, происходящих в какие-то, вообще говоря, случайные моменты времени. В БСУ мы имеем дело с различными видами потоков (напри-

мер, потоки задач, вызовов, справок в информационных системах; потоки отказов и восстановлений; потоки команд управления типа «включить», «отключить» в сложных иерархических системах управления рассредоточенными объектами; потоки требований на занятие определенного ресурса, причем в вычислительных системах — требование на занятие магистрали, внешнего запоминающего устройства, процессора, в системах связи — требование на занятие канала связи и т. д.).

При имитационном моделировании поток событий чаще всего воспроизводится через интервалы времени между соседними событиями. Если время между соседними событиями случайно, то в зависимости от вида распределения воспроизведение его в ЭВМ происходит в соответствии с теми способами, которые были рассмотрены при имитации непрерывных случайных величин, причем случайной величиной является длительность интервала между соседними событиями. Например, для простейшего потока событий время между событиями подчинено показательному закону; следовательно, имитация данного потока должна происходить в соответствии с выражением (6-5).

Модификация простейшего потока — поток Эрланга получается в результате имитации простейшего потока и последующего просенания его событий в соответствии с порядком этого потока. Регулярный поток в системе легко имитируется, так как он задается постоянным временем интервала между событиями. Аналогичным образом могут быть смоделированы и потоки более общего вида через задание соответствующего распределения интервалов между соседними событиями в потоке.

Рассмотренные выше способы имитации случайных факторов являются далеко не полным перечнем способов моделирования различных возможных случайных ситуаций, возникающих в БСУ; они наиболее характерны для выделенного класса систем.

Рассмотрим, как на их основе развывается имитация функционирования систем, динамика которых может быть отображена в классе марковских случайных процессов с дискретным числом состояний.

6-4. ИМИТАЦИЯ ФУНКЦИОНИРОВАНИЯ БСУ, ОПИСЫВАЕМЫХ В КЛАССЕ МАРКОВСКИХ ПРОЦЕССОВ

Из содержания гл. 4 было видно, что далеко не все системы, поведение которых может быть описано в классе марковских случайных процессов, могут быть изучены и исследованы аналитическим путем. Возникает вопрос, как подойти к моделированию марковских процессов.

Рассмотрим вопросы моделирования марковских процессов с конечным или счетным множеством состояний и дискретным временем перехода процесса из одного состояния в другое.

Пусть система, а следовательно, и процесс находятся в некотором множестве состояний Z ($z_i \in Z$, $i = 1, 2, \dots, n$). Будем считать для определенности, что n конечно. Динамика процесса будет считаться заданной, если задан граф перехода, а следовательно, и матрица перехода $\|P_{ij}\|$, где P_{ij} — вероятность перехода из состояния z_i в состояние z_j в некоторый дискретный момент времени $k\Delta t$ ($k = 1, 2, 3, \dots$). Переходы процесса из состояния в состояние могут осуществляться только в эти моменты времени. Считаем, что вероятности P_{ij} не зависят от времени, т. е. процесс является однородным марковским. Начальные

условия для процесса пусть заданы в виде распределения: $P_1(0), P_2(0), \dots, P_i(0), \dots, P_n(0)$, где $P_i(0)$ — вероятность того, что процесс при $t = 0$ находится в i -м состоянии.

Моделирование такого процесса основано на принципе имитации системы случайных событий: 1) сначала выбирается начальное число x_0 , задаваемое начальными вероятностями $P_1(0), P_2(0), \dots, P_n(0)$, и определяется номер начального состояния z_{m_0} , для которого оказывается справедливым условие $I_{m_0}^{(0)} \leq x_0 \leq I_{m_0+1}^{(0)}$; 2) затем выбирается следующее случайное число x_1 , которое также сравнивается с $I_r^{(1)}$, однако в качестве вероятностей P_i для определения границ $I_r^{(1)}$ используются элементы строки матрицы перехода с номером m_0 : $P_{m_0, 1}, P_{m_0, 2}, \dots, P_{m_0, n}$; путем сравнения устанавливается номер m_1 , для которого справедливо $I_{m_1}^{(1)} \leq x_1 \leq I_{m_1+1}^{(1)}$. Следовательно, следующим состоянием в данной реализации будет z_{m_1} . Таким образом, процесс повторяется далее.

В случае наличия установившегося режима влияния начальных вероятностей быстро уменьшается с ростом номера k . В результате моделирования получаем оценки вероятности $P_1, P_2, \dots, P_n$ нахождения процесса в соответствующих состояниях в установившемся режиме.

Рассмотрим, как будет выглядеть процесс моделирования, если переход из состояния в состояние может происходить в любой момент времени. В этом случае, как это уже было показано в гл. 3, однородный марковский процесс с конечным или счетным множеством состояний Z определяется: 1) распределением вероятностей состояний $P_i(t_0)$ в начальный момент t_0 ; 2) набором интенсивностей λ_{ij} переходов процесса из состояния z_i в состояние z_j , причем $\lambda_i = \sum_j \lambda_{ij}$ — интенсивность выхода из состояния i ($i \neq j$).

Один из возможных способов моделирования такого марковского процесса сводится к следующему. Пусть марковский процесс $z(t)$ удовлетворяет следующему соотношению: $z(t) = z_k$, $t_k \leq t < t_{k+1}$, где $t_0 < t_1 < t_2 < \dots < t_k < t_{k+1}$ — последовательность моментов изменения состояний, начиная от t_0 .

Для определения следующего состояния z_{k+1} и момента перехода в следующее состояние t_{k+1} генерируется ряд независимых случайных величин τ_{z_k, z_j} по показательному закону распределения с параметром λ_{z_k, z_j} [$z_j \in Z(z_k \rightarrow z_j)$]:

$$f_{z_k, z_j}(x) = \lambda_{z_k, z_j} \exp(-\lambda_{z_k, z_j} x), \quad x > 0$$

— и определяется $\tau_k = \min_{z_j} \tau_{z_k, z_j}$. Тогда следующий момент изменения состояния: $t_{k+1} = t_k + \tau_k$, а состояние, в которое переходит система в данный момент: $z(t) = z_{k+1}$, $t_{k+1} \leq t < t_{k+2}$, где z_{k+1} равно тому значению z_j , при котором τ_{z_k, z_j} является минимальным.

Такой способ моделирования марковского процесса называется развернутой рекуррентной имитацией. Физическая интерпретация такого подхода к моделированию марковских процессов с непрерывным временем позволяет непосредственно представить имитационные модели функционирования реальных систем. Предположим, что система в любой момент времени t может находиться в одном из состояний z_k , входящих в дискретное множество Z (конечное или счетное).

Если в момент t состояние системы есть z_k , то существует некоторое число физических причин, способных вывести ее из этого состояния ($z_k \rightarrow z_j$). Каждая из этих причин может вызвать переход системы во вполне определенное состояние, причем воздействие j -й причины (z_j соответствует состоянию, в которое может перейти система вследствие этой причины) проявляется через случайное время τ_{z_k, z_j} , распределенное по показательному закону с параметром λ_{z_k, z_j} , и не зависит от воздействия всех прочих причин. Решающее действие имеет та причина, которая вызовет случайное событие раньше других, т. е. процесс перейдет в то состояние z_j , для которого τ_{z_k, z_j} минимально. Для систем, представимых в виде моделей СМО, такими причинами являются: приход требования, окончание обслуживания, отказ обслуживающего прибора, восстановление обслуживающего прибора и т. д.

Программная имитация любых перечисленных выше случайных факторов делает имитационный эксперимент статистическим. Поэтому оценка интересующих нас показателей функционирования систем производится по многократному числу реализаций или по одной достаточно длинной реализации путем усреднения. Так как число реализаций (выборка имитационного эксперимента) влияет на точность получаемых оценок, рассмотрим далее некоторые основные принципы обоснования выбора числа реализаций.

6.5. ОПРЕДЕЛЕНИЕ ОБЪЕМА ИМИТАЦИОННЫХ ЭКСПЕРИМЕНТОВ

Объем эксперимента — это число реализаций, которое необходимо провести при имитационном моделировании, чтобы обеспечить требуемую статистическую точность результатов. При определении объема экспериментов обычно учитывается вид показателя эффективности. Показателем эффективности может быть вероятность выполнения той или иной задачи или некоторая скалярная функция параметров, алгоритма, структуры системы.

Рассмотрим первый случай. Сопоставим с фактом выполнения задачи некоторое событие A . Известно, что вероятность события A оценивается в процессе статистических испытаний как $P(A) = m/N$, где m — число случаев наступления события A при N реализациях. В силу предельной теоремы теории вероятностей частота m/N при достаточно большом N имеет распределение, близкое к нормальному, поэтому

$$P\left(\left|\frac{m}{N} - P\right| \leq t_{\alpha}\sigma\right) = \alpha. \quad (6-11)$$

Точность оценки m/N определяется следующей формулой:

$$\varepsilon = t_{\alpha}\sigma, \quad (6-12)$$

где α — доверительная вероятность; t_{α} — квантиль нормального закона, соответствующий заданному значению α и определяемый по таблицам нормального распределения [8]:

$$t_{\alpha} = \Phi^{-1}\left(\frac{\alpha}{2}\right); \quad \Phi(t_{\alpha}) = \frac{1}{\sqrt{2\pi}} \int_0^{t_{\alpha}} \exp\left(-\frac{z^2}{2}\right) dz = \alpha/2,$$

где Φ — функция Лапласа; Φ^{-1} — функция, обратная функции Лапласа; σ^2 — дисперсия частоты m/N , определяемая по формуле:

$$\sigma^2 = \frac{P(1-P)}{N}. \quad (6-13)$$

Из формул (6-12) и (6-13) можно получить соотношение для определения необходимого числа реализаций при заданной точности оценки ε и доверительной вероятности α :

$$N = t_{\alpha}^2 \frac{P(1-P)}{\varepsilon^2}. \quad (6-14)$$

Для $\alpha = 0,95$ и при различных значениях P и ε число реализаций N можно определить по табл. 6-1.

Из табл. 6-1 видно, как значительно растет число имитационных экспериментов в случае повышения требований к точности получаемых результатов. Особенно остро проблема выбора числа имитационных экспериментов встает при моделировании редких событий, возникающих в системе с малыми вероятностями P .

Таблица 6-1

Перейдем в выражении (6-13) от абсолютной точности ε к относительной точности $\delta = \varepsilon/P$; тогда формула (6-13) примет следующий вид:

$$N = \frac{t_{\alpha}^2 (1-P)}{P\delta^2} \approx \frac{t_{\alpha}^2}{P\delta^2},$$

откуда видно, что при малых P очень значительно возрастает N .

Данный факт сужает область применения метода имитационного моделирования и даже исключает его использование без специальных способов повышения статистической точности результатов при малом числе экспериментов [41]. Подобные задачи возникают при имитационном моделировании поведения БСУ с учетом их неидеальной работы, т. е. при наличии сбоев и отказов.

Аналогично решаются вопросы определения числа экспериментов и во втором случае, когда оценивается некоторый показатель эффективности E , являющийся функцией параметров, алгоритма функционирования, структуры системы. Оценка среднего значения показателя E по множеству реализаций N определяется как

$$E = \frac{1}{N} \sum_{i=1}^N E_i,$$

где E_i — значение показателя эффективности в i -м эксперименте.

По центральной предельной теореме при большом N среднее арифметическое E имеет распределение, близкое к нормальному с математическим ожиданием m_E и дисперсией σ_E^2/N , где σ_E^2 — дисперсия оцениваемой случайной величины. Поэтому аналогично (6-11) имеем

$$P(|E - m_E| \leq t_{\alpha}\sigma_E/\sqrt{N}) = \alpha,$$

$$N = \frac{t_{\alpha}^2 \sigma_E^2}{\varepsilon^2} \quad (6-15)$$

Как видно из (6-14), (6-15), для определения объема имитационных экспериментов необходимо знать значения P либо σ_E , а они, как правило, обычно бывают неизвестны. В этом случае либо проводят предварительную «пристрелку» для определения оценок P и σ_E , либо используют последовательный алгоритм определения необходимого объема имитационных экспериментов [41].

Вначале для проведения имитационных экспериментов выделяется некоторое число прогонов ΔN и вычисляется оценка дисперсии s^2 для интересующего показателя эффективности E . Если оказывается, что $t_{\alpha}^2 s^2 > \varepsilon^2$, где α — требуемая доверительная вероятность, а ε — требуемая точность получения оценки, то выделяется еще дополнительное число прогонов ΔN , после которых производится уточнение оценки E . Таким образом процесс повторяется далее до получения требуемой статистической точности при условии, что число экспериментов не превышает максимально допустимого $N_{\max}$.

Аналогично строится последовательный алгоритм и для случая, когда необходимо произвести оценку вероятности какого-либо события. Во многих случаях вопрос об априорном определении необходимого числа экспериментов остается открытым и лучше, чем последовательный алгоритм уточнения объема выборки, вряд ли что можно предложить. Практика исследования БСУ показывает, что для многих встречающихся случаев процесс функционирования обладает свойствами стационарности и эргодичности. Это означает, что вероятностные характеристики процесса в установившемся режиме не зависят от времени, причем предельный режим, установившийся в системе через некоторое время ее работы, не зависит от того, какие были начальные условия и первоначальный период работы системы. Из этого можно сделать важный для практического исследования вывод: если речь идет об исследовании не начального, переходного периода, а установившегося режима работы системы, то о свойствах системы при имитационном моделировании можно судить по одной достаточно длинной реализации [41]. При этом интересующие нас показатели эффективности находятся как средние по времени для одной реализации. Точное определение длины реализации T по заданной доверительной вероятности α и допустимой ошибке ε представляет собой достаточно сложную задачу.

Приближенно с достаточной для практических целей точностью длину реализации T можно определить, пользуясь следующим соотношением:

$$T \approx NT_0,$$

где N — число реализаций длительностью T_0 , которое определяется в соответствии с формулами (6-14), (6-15); T_0 — наблюдаемый период функционирования системы.

Для обработки и анализа данных, полученных в результате моделирования, применяются традиционные методы математической статистики. В ряде случаев алгоритмы обработки результатов имитационного эксперимента оформляются как отдельные подпрограммы и включаются в общую имитационную модель. Результаты исследования обычно приводятся в виде табличных или графических зависимостей, кото-

рые позволяют дать ответ на те вопросы, которые были сформулированы в начале исследования. На рис. 6-4 приведена схема, иллюстрирующая последовательность выполнения основных работ при имитационном исследовании на ЭВМ различных процессов. При составлении программы для имитационных моделей могут быть использованы как универсальные алгоритмические языки программирования типов АЛГОЛ, ФОРТРАН, так и специализированные языки программирования типов СИМУЛА, СЛЭНГ и др. Универсальные языки хорошо приспособлены лишь для описания вычислительных операций, однако служебный механизм имитационного моделирования должен создаваться автором модели. Кроме того, существенным недостатком программирования на этих языках является отсутствие единого методологического подхода к описанию БСУ.

Специализированные языки позволяют: 1) создать единый подход к анализу и описанию различных систем; 2) обеспечить возможность формального описания систем и автоматической трансляции такого описания в коды вычислительной машины; 3) отображать в программной модели факторы, присущие широкому классу систем с дискретными событиями; 4) обеспечить удобочитаемость описаний системы и простоту обмена информацией между моделью и исследователем. Поэтому созданию специализированных языков в настоящее время уделяется большое внимание [56].

Рассмотренные в данной главе вопросы не являются полным исчерпывающим изложением всех аспектов имитационного моделирования как инструмента для определения характеристик БСУ, а скорее являются только введением в интересную интенсивно развивающуюся область имитационных исследований.

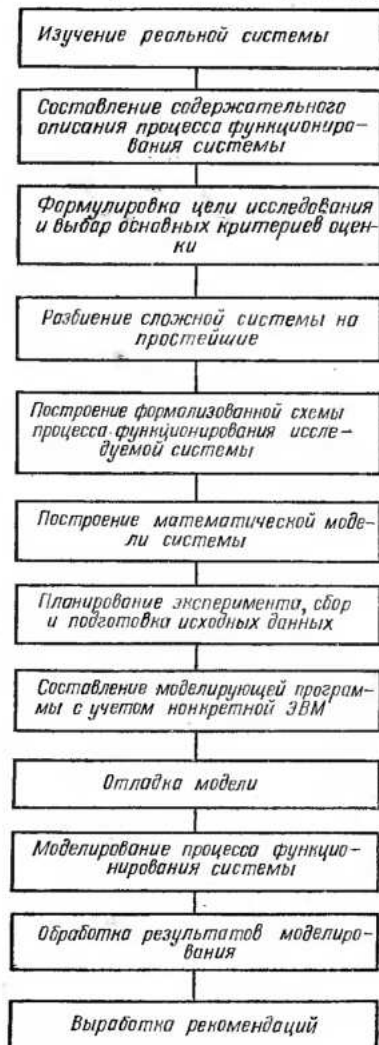


Рис. 6-4. Основные этапы имитационного моделирования

ИНФОРМАЦИОННЫЙ ПОДХОД К АНАЛИЗУ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

7-1. ИНФОРМАЦИОННЫЕ ЦЕПИ. УСТАНОВИВШИЙСЯ РЕЖИМ

Любые системы управления, из каких бы элементов они ни слагались и какие бы цели ни преследовали, по существу являются системами передачи и переработки информации, поэтому ниже излагаются методы информационного анализа цепей управления.

Источники и потребители информации. Под информационной цепью мы понимаем совокупность взаимодействующих источников, преобразователей и потребителей информации. С практической точки зрения наибольший интерес представляют информационные цепи систем управления, понимаемые в самом широком смысле как целеустремленные биологические, социальные системы и автоматы. Например, информационная цепь системы управления дыханием высших животных состоит из: нервного центра управления дыханием, выполняющего функции источника информации; нервных волокон, передающих управляющую информацию легочной мускулатуре; самой этой мускулатуры, выступающей в роли потребителя управляющей информации (информационной нагрузки); потоков крови, доставляющих в центр управления информацию о насыщенности организма кислородом и выполняющих функции обратной связи [16].

Примером информационной цепи социальной системы управления может служить структура государственного управления, где законодательные органы выступают в качестве источников управляющей информации для исполнительной власти, выполняющей в этом случае роль информационной нагрузки, а отчеты местных органов, газетные репортажи, статьи и реакция избирателей обеспечивают обратный поток информации к законодательным органам.

Информационная цепь автоматической системы управления состоит из регулятора, служащего источником управляющей информации для объекта управления, который выступает в роли информационной нагрузки, и устройств обратной связи, обеспечивающих поток информации от объекта управления к регулятору.

В общем случае информационные цепи представляют собой довольно сложные сильно разветвленные многоконтурные иерархические структуры, подобные системам государственного управления или системам управления жизнедеятельностью биологических образований. Однако в простейшем случае элементарная информационная цепь состоит из одного источника информации (рис. 7-1), одного потребителя информации (информационной нагрузки) и связывающих их проводников информации (информационных каналов).

Такие информационные цепи всегда замкнуты на источник либо посредством каналов прямой и обратной связи, либо (в отсутствие одного из них) посредством логических связей между источником и нагрузкой. Так, в отсутствие обратной связи в некоторых системах программного управления ее роль выполняет уверенность в предопределенности поведения объекта под воздействием программы. Напротив, в чисто познавательных (измерительных) системах, где отсутствует канал воздействия на объект познания (прямая связь), его функции компенсируются установлением логических связей между познающим объектом и объектом познания по мере изучения последнего. Наконец, в отсутствие, казалось бы, заметной связи между источником и приемником связь между ними обеспечивает интуиция.



Рис. 7-1. Структура информационной цепи

Состояние окружающей нас материи характеризуется некоторой неопределенностью, или энтропией, $H_0 = -\log P_0$, которая выступает в роли информационного потенциала (сути) события, априорная вероятность которого равна P_0 . Целью и смыслом всякого управления является изменение в ту или иную сторону этой априорной вероятности события до некоторого нового значения $P_{усл}$, которому соответствует новое значение потенциала $H_{усл} = -\log P_{усл}$, где $P_{усл}$ — вероятность события при условии управления им.

Таким образом, сущность управления, осуществляемого источником информации, может быть охарактеризована некоторым информационным напряжением

$$\Delta H = H_0 - H_{\text{усл}} = \log \frac{P_{\text{усл}}}{P_0}. \quad (7-1)$$

В управленческой деятельности, представляющей наибольший прикладной интерес, источником информации являются обычно люди, коллективы людей либо технические устройства. Так, оператор, управляющий прокатным станком, является источником управляющей информации, напряжение которого равно логарифму отношения вероятности успешной работы оператора к вероятности успешной работы стана при устранении оператора. Точно так же министерство является источником управляющей информации для отрасли, имея информационное напряжение, определяемое вероятностями выполнения отраслью государственного плана при наличии и устранении всего управленческого аппарата министерства. Наконец, система программного управления станком является источником информации, напряжение которого определяется вероятностями успешного выполнения станком нужной операции при наличии и отсутствии соответствующей программы.

Из выражения (7-1) можно сделать вывод, что информационное напряжение (суть) источника ΔH может быть как положительным, когда его целью является увеличение вероятности события, так и отрицательным, когда его целью является снижение вероятности события. Если же $P_{\text{усл}} = P_0$, то напряжение источника равно нулю, т. е. его роль в управлении несущественна и он не имеет смысла.

Информационное напряжение, как и энтропию, можно измерять в различных единицах в зависимости от выбора основания логарифма в (7-1) или от склонности автора к тем или иным интерпретациям. Нередко [2] информационную энтропию отождествляют с термодинамической энтропией, измеряя их в одинаковых единицах; однако мы в дальнейшем будем пользоваться в качестве единиц напряжения б и т а м и, получающимися, если в (7-1) использовать двончные логарифмы, поскольку это обеспечивает хорошо интерпретируемую размерность всех других информационных величин.

Пример 7-1. Подсчитаем информационное напряжение, которым обладает система оповещения противовоздушной обороны в такой ситуации, когда к очень важному объекту, непосредственно обороняемому зенитно-ракетной установкой, сквозь внешнее оборонительное кольцо прорвались два однотипных самолета противника, один из которых потратил свой наиболее разрушительный боезапас в процессе прорыва. В этом случае, не имея уже времени поразить оба самолета,

зенитно-ракетная установка, не связанная с системой оповещения, поразит самолет с нерастраченным боезапасом с вероятностью 0,5. Если же система оповещения успеет правильно ориентировать установку, то та поразит самолет, обладающий разрушительной мощностью, скажем, с вероятностью 1. Тогда согласно (7-1)

$$\Delta H = \log 2 = 1 \text{ бит.}$$

Стало быть, в описанной ситуации напряжение источника информации (системы оповещения) составляет 1 бит.

Информационное напряжение не следует путать с информацией, которую нередко тоже измеряют в битах. Мы вернемся к вопросу об информации в свое время, пока же отметим, что в рассмотренном примере источник, имеющий напряжение 1 бит, способен передать сколько угодно информации объекту управления (зенитно-ракетной установке) в зависимости от времени, которым он будет располагать.

Информационное напряжение источника может изменяться во времени свое значение или знак, если важность достижения цели неодинакова в различные моменты времени. Если это изменение происходит периодически с постоянным периодом T , как при работе автоматических систем в режиме автоколебаний, то такое переменное напряжение удобно характеризовать действующим значением

$$\Delta H_d = \sqrt{\frac{1}{T} \int_0^T [\Delta H(t)]^2 dt} = \sqrt{\frac{1}{T} \int_0^T \left[\log \left(\frac{P_{\text{усл}}}{P_0} \right) \right]^2 dt} = \sigma(\Delta H), \quad (7-2)$$

которое есть не что иное, как среднеквадратическое напряжение $\sigma(\Delta H)$. Наконец, при случайных изменениях сути сигнала x

$$\Delta H_0 = \int_{-\infty}^{\infty} f(x) \Delta H dx; \quad \Delta H_d^2 = \int_{-\infty}^{\infty} f(x) \Delta H^2 dx,$$

где ΔH_0 и ΔH_d — среднее и действующее значения сущности сигнала; $f(x)$ — плотность распределения вероятности x .

Если же $\Delta H = A \sin(t/T)$, то согласно (7-2) действующее значение переменного напряжения составит $\Delta H_d = A/\sqrt{2}$, что почти в 1,5 раза меньше максимального мгновенного значения напряжения.

Выданная источником информация, как отмечалось выше, поступает к исполнительным органам и системам,

которые являются информационной нагрузкой источника, а затем возвращается в источник в форме информации обратной связи. Если исполнительная система не обладает ни памятью, ни привычками, то единственной ее характеристикой в рассматриваемом аспекте является информационное сопротивление, т.е. время ее реакции на полученную информацию (время исполнения), которое исчисляется от момента выхода управляющей информации из источника до момента получения источником сигнала обратной связи о достижении поставленной

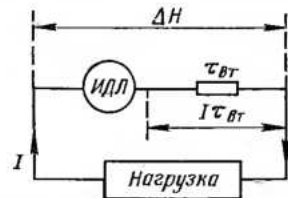


Рис. 7-2. Простейшая информационная цепь

цели. Так, в чисто человеческих системах управления, где распоряжения отдаются в устной форме непосредственно или по телефону, информационное сопротивление системы исполнения равно времени исполнения от момента, когда распоряжение сформулировано, до момента, когда поступил доклад об исполнении. При этом время, необходимое для принятия самого решения, для его изложения и для осмысленного восприятия доклада об исполнении, является внутренним информационным сопротивлением источника информации, обратным его пропускной способности $I_{\max}$.

Таким образом, для системы без привычек и памяти (рис. 7-2) имеет место информационный закон Ома

$$I = \Delta H / \tau_{\text{н}}, \quad (7-3)$$

где $\tau_{\text{н}} = \tau - \tau_{\text{вт}}$ — информационное сопротивление нагрузки; τ и $\tau_{\text{вт}}$ — информационные сопротивления соответственно всей цепи и источника (внутреннее); I — информационный ток в цепи нагрузки.

Из (7-3) следует, что при однократном достижении цели сквозь систему проходит информация $J_{\text{ц}}$, численно равная напряжению источника:

$$J_{\text{ц}} = I \tau_{\text{н}} = \Delta H. \quad (7-4)$$

При длительной работе системы в течение времени T сквозь нее протекает информация

$$J = \int_0^T I dt = \int_0^T \frac{\Delta H}{\tau} dt. \quad (7-5)$$

Например, если при посадке самолета в его салоне зажигается табло «Застегнуть ремни» и горит в течение времени T , а для исполнения команды нужно время τ , то пассажиры успевают получить информацию

$$J = -\frac{T}{\tau} \log P,$$

где P — вероятность самопроизвольного исполнения команды при внезапно испортившемся табло и отсутствии иных средств доведения команды до дисциплинированных ($P_{\text{ysl}} = 1$) пассажиров.

Формула (7-5) справедлива, конечно, только в тех случаях, когда напряжение и ток не меняют знака, как при рассмотрении светящегося табло в салоне самолета. В тех же весьма частых случаях, когда напряжение и ток периодически изменяют знак, в формуле (7-5) следует подставлять действующее значение тока, которое может быть определено следующим образом:

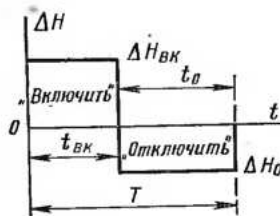


Рис. 7-3. Знакопеременная энтропия

$$I_{\text{д}} = \sqrt{\frac{1}{T} \int_0^T i^2 dt} = \sigma(i), \quad (7-6)$$

где i — мгновенное значение информационного тока; T — период изменения тока.

Например, если передается последовательность команд «включить», «отключить» с периодом T и продолжительностью соответственно $t_{\text{в}}$ и $t_{\text{о}}$ (рис. 7-3), то соответствующие им положительное и отрицательное информационные напряжения $\Delta H_{\text{в}}$ и ΔH_0 определяются соответствующими априорными вероятностями этих состояний объекта управления до подачи команд. При этом, если время включения (информационное сопротивление) объекта составляет $\tau_{\text{в}}$, а время отключения $\tau_{\text{о}}$, то вышеприведенным напряжениям отвечают токи $i_{\text{в}} = \Delta H_{\text{в}} / \tau_{\text{в}}$ в течение $t_{\text{в}}$ и $i_{\text{о}} = \Delta H_0 / \tau_{\text{о}}$ в течение $t_{\text{о}}$.

Воспользовавшись (7-6), получим действующее значение тока:

$$I_{\text{д}} = \sqrt{\frac{1}{T} \left(\frac{t_{\text{в}} \Delta H_{\text{в}}^2}{\tau_{\text{в}}^2} + \frac{t_{\text{о}} \Delta H_0^2}{\tau_{\text{о}}^2} \right)} = \sigma(i). \quad (7-7)$$

В важном частном случае, когда оба состояния объекта управления априорно равновероятны, а $t_n = t_o = T/2$ и $\tau_n = \tau_o = \tau$, имеем (бит/с)

$$I_d = i_n = \text{mod } i_o = 1/\tau.$$

Если же при прочих равных условиях ΔH изменяется по синусоидальному закону $\Delta H = B \sin(t/T)$, то действующее значение тока составит (бит/с) $I_d = \frac{B}{\tau \sqrt{2}}$.

Естественно, что эффективность источника зависит от того, насколько быстро он выдает управляющую информацию при изменении состояния нагрузки. Запаздывание, имеющееся в источнике, обесценивает выданную им управляющую информацию и выполняет функции внутреннего информационного сопротивления источника.

Действительно, если при стрельбе по быстро движущимся целям источник управляющей информации медленно вычисляет координаты целей, то может случиться, что, когда он их выдаст, цель окажется уже вне пределов досягаемости средств поражения, иначе говоря, напряжение такого источника упадет практически до нуля и он будет бесполезен, хотя чисто теоретически, без учета запаздывания, он вычислял бы координаты целей со сверхвысокой точностью и на холостом ходу в отсутствие целей обладал бы весьма высоким напряжением. Это напряжение источника, которым он обладает на холостом ходу без информационной нагрузки, иными словами, без учета внутреннего сопротивления, мы будем называть информационно-движущей логикой (ИДЛ) источника. При наличии нагрузки информационный ток I создает падение напряжения на внутреннем сопротивлении $\tau_{вт}$ (см. рис. 7-2), которое снижает ИДЛ до рабочего напряжения на величину $I\tau_{вт}$, так что

$$\Delta H = h - I\tau_{вт}, \quad (7-8)$$

где h — ИДЛ источника.

Таким образом, чем больше запаздывание в источнике, т. е. чем больше времени занимает процесс переработки информации и принятия решения, тем согласно (7-8) меньше его напряжение по сравнению с ИДЛ, а значит, тем меньше он способен изменить вероятность достижения цели управления. Это свойство более усиливается по мере увеличения нагрузки, т. е. по мере роста информационного тока. Поэтому при проектировании источника для работы на опре-

деленную нагрузку (на заданный информационный ток) приходится с учетом его внутреннего сопротивления завышать его ИДЛ на $I\tau_{вт}$ с целью обеспечить заданную вероятность нужного события.

Переписав (7-8) в форме

$$P_{усл} = P_h \exp(-I\tau_{вт} \ln 2), \quad (7-9)$$

получим, что для больших информационных токов источники с заметным внутренним сопротивлением (запаздыванием) $\tau_{вт}$ могут обеспечить лишь сравнительно низкую вероятность $P_{усл}$ желаемого события. Поскольку любые реальные источники информации (люди и ЭВМ) обладают конечным информационным сопротивлением, мы рассмотрим способы его уменьшения посредством схемных ухищрений; пока же отметим, что применительно к человеку ИДЛ характеризует его потенциальные творческие возможности при практически неограниченном времени, отведенном для принятия решений.

Что касается внутреннего информационного сопротивления человека, то оно характеризует только быстроту соображения, например скорость арифметических операций, вне зависимости от потенциальных возможностей индивида. Информационное же напряжение человека как источника информации определяется согласно (7-8) совокупным действием факторов. В результате часто весьма одаренные, но с замедленной реакцией люди оказываются беспомощными при оперативном управлении быстроизменяющимися ситуациями, обеспечивая согласно (7-9) лишь сравнительно низкую вероятность достижения цели управления. Напротив, люди даже весьма ограниченные, но решительные и с хорошей реакцией, обладая низким информационным сопротивлением, способны при ограниченных значениях информационных токов обеспечить в соответствии с (7-9) довольно успешное оперативное управление.

Однако положение изменяется радикальным образом, когда речь заходит о стратегическом планировании, на которое отводится достаточно времени, чтобы информационные токи были малы, а потери напряжения $I\tau_{вт}$ даже при значительных $\tau_{вт}$ оставались несущественными. В этом случае согласно (7-8) $\Delta H \approx h$ и $P_{усл} \approx P_h$.

Таким образом, здесь быстрота соображения не играет существенной роли, а высокая вероятность достижения цели обеспечивается только ИДЛ, т. е. только талантом и опытом человека. Эти соображения следует принимать

во внимание при распределении кадров по ступеням управленческой иерархии, на вершине которой принимаются относительно нечастые, зато весьма ответственные решения, а на нижних ступенях, напротив, принимаются не слишком важные решения, зато в большом числе и в ограниченное время.

Соединения источников и приемников информации. В процессе работы часто приходится комбинировать элементарные источники информации с целью увеличения ИДЛ или их пропускной способности. При этом применяют соответственно последовательное или параллельное соединение элементарных источников. При последовательном соединении (рис. 7-4, а) через источники и нагрузку течет один

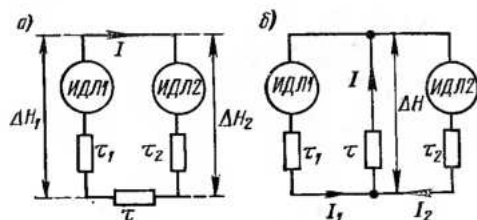


Рис. 7-4. Соединения источников информации: а — последовательное; б — параллельное

и тот же информационный ток I , а их результирующее информационное напряжение равно сумме напряжений отдельных источников $\Delta H = \Delta H_1 + \Delta H_2$, что увеличивает вероятность достижения желаемого события, так как

$$\log(P/P_0) = \log(P_1/P_0) + \log(P_2/P_0) = \log(P_1 P_2 / P_0^2),$$

где $P_1 < P_1 P_2 / P_0 > P_2$; $P = P_1 P_2 / P_0$.

Вместе с тем последовательное соединение источников приводит к увеличению их суммарного внутреннего сопротивления $\tau_{\text{вт}} = \tau_1 + \tau_2$, поэтому оно эффективно только в том случае, если суммарная задержка решения в источниках значительно меньше информационного сопротивления нагрузки τ , представляющего собой время реакции исполнительного органа (время ощутимых изменений управляемого процесса: $\tau \gg \tau_{\text{вт}}$). В этом случае $I = (h_1 + h_2) / \tau$; в противном случае, т. е. при $\tau_{\text{вт}} \gg \tau$, последовательное соединение источников ничего не дает, так как информационный ток не изменяется, что хорошо видно при одина-

ковых источниках:

$$I = \frac{h_1 + h_2}{\tau_1 + \tau_2} = \frac{h_1}{\tau_1}.$$

Это обстоятельство нужно учитывать при организации источников, в которых отделы (люди или автоматы) последовательно во времени осуществляют экспертизу (визируют) с разных точек зрения подготовленное одним из них решение, что, с одной стороны, повышает ИДЛ источника (вероятность достижения цели управления), но, с другой стороны, в случае длительной бюрократической волокиты в вышестоящих контрольных органах ($\tau_{\text{вт}} \gg \tau$), может сделать бессмысленным существование последних, так как они, не увеличивая информационный ток, экономически обременительны.

В общем случае, т. е. при соизмеримых информационных сопротивлениях источников и нагрузки, с учетом (7-8) имеет место следующее основное выражение:

$$I = \frac{h_1 + h_2}{\tau + \tau_1 + \tau_2} = \frac{\Delta H_1 + \Delta H_2}{\tau}. \quad (7-10)$$

В случае параллельного соединения (рис. 7-4, б) источники работают с одинаковым напряжением ΔH , а в нагрузку поступает их суммарный ток $I = I_1 + I_2$.

Это имеет место, когда, например, главки одного министерства параллельно обеспечивают выполнение различных частей единого общего для министерства плана отрасли или когда жизнь производственного предприятия направляется органами отраслевого управления, местными партийными и профсоюзными органами, работающими как параллельные источники управляющей информации на одну нагрузку. В результате этого эквивалентное внутреннее сопротивление источников значительно снижается, т. е. практически сокращается время выработки ими управляющих решений по сравнению со случаем, когда всю разнородную управляющую информацию должен выдать лишь один из источников.

Действительно, если $h_1 = h_2 = h$, то имеет место соотношение $I_1 \tau_1 = I_2 \tau_2 = I \tau_{\text{вт}}$, где $I = I_1 + I_2$, а $\tau_{\text{вт}}$ — эквивалентное внутреннее сопротивление источников.

Из этого соотношения, исключая токи, нетрудно получить

$$\tau_{\text{вт}} = \frac{\tau_1 \tau_2}{\tau_1 + \tau_2} \quad \text{или} \quad \frac{1}{\tau_{\text{вт}}} = \frac{1}{\tau_1} + \frac{1}{\tau_2}. \quad (7-11)$$

Из (7-11) следует, что эквивалентное внутреннее сопротивление параллельных источников с одинаковыми ИДЛ меньше внутреннего сопротивления любого из них и что их эквивалентная проводимость (величина, обратная сопротивлению) равняется сумме проводимостей каждого из них.

При этом ток через нагрузку составляет

$$I = \frac{h}{\tau_{вт} + \tau}.$$

Это явно больше, чем в состоянии обеспечить один источник при той же нагрузке τ :

$$I_1 = \frac{h}{\tau_1 + \tau} < I.$$

Однако этот эффект имеет место, если внутренние сопротивления источников соизмеримы с сопротивлением нагрузки. В противном случае, когда $\tau \gg \tau_{вт}$, параллельные источники дают такой же ток, как и каждый из них в отдельности.

Отметим, что, как это следует из схемы (рис. 7-4, б), параллельное соединение источников только тогда выполняет свою роль, когда источники имеют равные ИДЛ, т. е. равнокомпетентны в деле управления. В противном случае тот из них, компетентность которого ниже других, сам становится для них нагрузкой и потребляет (вместо того, чтобы выдавать) управляющую информацию, не только шунтируя основную нагрузку, но и перегружая остальные источники управляющей информации. Такое положение нередко складывается, когда в одной из сфер параллельного управления оказываются люди, не имеющие соответствующей профессиональной подготовки.

Рассмотрим, например, случай управления энергосистемой двумя или большим числом равноправных параллельно работающих диспетчеров. Если все они равнокомпетентны, т. е. если вмешательство любого из них увеличивает вероятность успешной работы на одну и ту же величину, то их совместная параллельная работа ускоряет управление, уменьшая согласно (7-11) внутреннее сопротивление диспетчерского пункта во столько раз, сколько там диспетчеров. Если же один из них имеет низкую квалификацию или недостаточный опыт, т. е. если его вмешательство увеличивает вероятность успешного управления в меньшей степени, чем вмешательство каждого из остальных диспетчеров, то остальные вынуждены взять на себя часть его

работы, что приводит к снижению их потенциала до уровня низкоквалифицированного диспетчера, который в этой ситуации выступает в роли дополнительной нагрузки.

Действительно, для источников с различными ИДЛ имеем $h_1 - I_1\tau_1 = h_2 + I_2\tau_2 = I\tau$ при условии, что $h_1 > h_2$. При этом условии ток I_2 течет не из источника, а в источник ИДЛ2 (h_2) как в нагрузку. Из приведенных соотношений с учетом $I = I_1 - I_2$ имеем

$$\left. \begin{aligned} I &= \frac{h_1\tau_2 + h_2\tau_1}{\tau(\tau_1 + \tau_2) + \tau_1\tau_2}; \\ I &= \frac{h_1\tau_2 + h_2\tau_1}{\tau(\tau_1 + \tau_2)}, \end{aligned} \right\} \quad (7-12)$$

если $\tau \gg \tau_1 + \tau_2$.

Выражения (7-12) доказывают, что ток через нагрузку при параллельной работе на нее источников с различными ИДЛ всегда меньше, чем ток при работе на ту же нагрузку только одного источника с большей ИДЛ (h_1). Кроме того, этот источник, работая на большую нагрузку в паре с другим, перегружается последним, так как вынужден генерировать ток

$$I_1 = \frac{h_1 - h_2}{\tau_1 + \tau_2},$$

который при всех обстоятельствах больше тока от этого источника в отсутствие параллельного ему:

$$I'_1 = \frac{h_1}{\tau + \tau_1} < I_1.$$

Наконец, можно подсчитать, что параллельный источник снижает первоначальное напряжение источника с большей ИДЛ (h_1) по сравнению с работой его только на нагрузку на

$$\Delta H = (I_1 - I'_1)\tau_1 = (h_1 - h_2) \frac{\tau_1}{\tau_1 + \tau_2}.$$

Если $\tau_1 = \tau_2$, то $\Delta H = (h_1 - h_2)/2$, т. е. уменьшение большего напряжения составит половину разницы ИДЛ источников.

Таким образом, параллельная работа различных источников допустима лишь при условии их равной компетентности в вопросах управления, что необязательно подразумевает одинаковость их образования или положения в социальной структуре. Например, составитель программы для станка с программным управлением и разнорабочий,

очищающий станок от стружки, окажутся равнокомпентными, т. е. имеющими одинаковые ИДЛ, если вероятность сбоев в работе станка из-за ошибок программы окажется равной вероятности сбоев из-за попадания стружки в систему управления.

Нетрудно себе представить, что не только источники, но и информационные сопротивления (нагрузки) могут быть соединены как последовательно (рис. 7-5, а), так и параллельно (рис. 7-5, б); при этом под последовательным соединением нагрузок понимается не схемная, а временная последовательность поступления информации в нагрузки, что характерно для временной селекции. О последовательном соединении можно говорить только тогда, когда информация от источника либо сначала поступает в первую нагрузку, и лишь после выполнения программы — во вторую, либо когда одна из нагрузок после выполнения своей программы сама передает ее другой нагрузке.

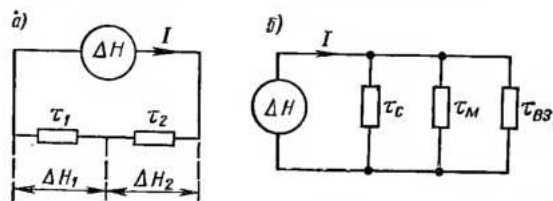


Рис. 7-5. Соединения нагрузок: а — последовательное; б — параллельное

Так или иначе, общее время работы последовательно соединенных нагрузок должно равняться сумме времен отработки своих программ отдельными нагрузками, так что для схемы (рис. 7-5, а) $\tau = \tau_1 + \tau_2$. Иными словами, при последовательном соединении нагрузок их информационные сопротивления должны складываться. При этом напряжение источника распределяется между нагрузками пропорционально их сопротивлениям. Соотношения, характеризующие схему рис. 7-5, а, следующие:

$$\left. \begin{aligned} \Delta H_1 &= I\tau_1; \quad \Delta H_2 = I\tau_2; \quad \Delta H = \Delta H_1 + \Delta H_2; \\ I &= \Delta H/\tau; \quad \Delta H_1/\Delta H = \tau_1/\tau; \quad \Delta H_2/\Delta H = \tau_2/\tau. \end{aligned} \right\} \quad (7-13)$$

Например, если для выполнения поставленной цели — разгрома вражеской группировки — наступающим вдоль железной дороги войскам необходимо захватить последо-

вательно расположенные на этой дороге опорные пункты А, Б и В, то, получив приказ о взятии пункта А, войска затратят на его штурм время τ_A , прежде чем получат приказ о взятии пункта Б, на что тоже потребуется время τ_B .

Наконец, получив приказ о взятии пункта В и затратив на его штурм время τ_B , войска выполнят задачу, израсходовав общее время $\tau_A + \tau_B + \tau_B$. При этом очевидно, что чем важнее опорный пункт, тем лучше он укреплен и тем больше времени потребуются для его штурма, но зато и тем больший вклад его успешное взятие вносит в дело выполнения всей задачи по разгрому группировки врага, т. е. тем сильнее изменение вероятности достижения цели и соответственно — информационного напряжения. В частности, если все опорные пункты равнозначны и на их штурм затрачивается одинаковое время τ , а вероятность того, что, не получая приказов и не зная цели наступления, войска сами займут нужный опорный пункт, составляет P , то легко сосчитать, что отношение информационного сопротивления τ к общему информационному сопротивлению 3τ составит $1/3$, а отношение изменения напряжения при штурме каждого пункта $\Delta H_k = -\log P$ к напряжению всей операции $\Delta H = -\log P^3 = -3 \log P$ также равно $1/3$, что, с одной стороны, подтверждает соотношения (7-13), а с другой стороны, требует от командования согласно (7-4) в три раза большего запаса информации, чем при штурме одного опорного пункта.

В случае параллельного соединения нагрузок, т. е. нескольких исполнительных органов, каждый из которых в состоянии в одиночку достичь цели, они одновременно получают управляющую информацию и одновременно исполняют операции, необходимые для достижения цели (рис. 7-5, б), соответственно сокращая время ее достижения.

Так, если войскам, подступившим к морскому порту с суши, для захвата порта необходимо время τ_c , морскому десанту для этого требуется время τ_m , а воздушному десанту — время $\tau_{вз}$, то все вместе они возьмут порт значительно скорее, чем каждый в отдельности.

Действительно, поскольку информационный потенциал порта

$$\Delta H = -\log P,$$

где P — вероятность захвата без приказа, не зависит от рода штурмующих войск, то очевидно имеет место соот-

$$\Delta H = I_c \tau_c = I_m \tau_m = I_{вз} \tau_{вз},$$

из которого следует, что информационные токи, связанные с тем или иным родом войск, обратно пропорциональны их информационным сопротивлениям:

$$I_c/I_m = \tau_m/\tau_c; \quad I_m/I_{вз} = \tau_{вз}/\tau_m; \quad I_c/I_{вз} = \tau_{вз}/\tau_c. \quad (7-14)$$

Иными словами, их эквивалентная информационная проводимость $1/\tau$, обратная эквивалентному информационному сопротивлению τ параллельного соединения информационных нагрузок, равна сумме информационных проводимостей отдельных нагрузок:

$$1/\tau = 1/\tau_c + 1/\tau_m + 1/\tau_{вз}. \quad (7-14a)$$

Естественно, при этом источник должен обеспечить ток, равный сумме токов отдельных нагрузок $I = I_c + I_m + I_{вз}$, т. е. управление одновременным наступлением нескольких родов войск требует от командования переработки за меньшее время такого же объема информации, как при наступлении одного рода войск.

В качестве параллельных основной нагрузке могут выступать посторонние нагрузки, создающие утечки информации. В такой роли нередко оказываются всевозможные контролирующие органы, которые иногда настолько перегружают источник требованием от него разных справок и отчетов, что он снижает свое напряжение, т. е. снижает вероятность достижения своей прямой цели, поскольку не в состоянии параллельно обеспечивать информацией объект управления и контрольные органы. Конечно, когда такая параллельная работа запланирована заранее, т. е. когда источник выбран в расчете на такую нагрузку, а контролирующие органы не превышают своих полномочий, напряжение источника соответствует требуемой вероятности достижения цели (надежности управления), но за счет выработки им избыточного (не необходимого для управления) информационного тока $I_{ц}$. Это можно обеспечить, либо увеличив напряжение источника на $\Delta H = I_{ц} \tau_{вт}$, где $\tau_{вт}$ — внутреннее сопротивление (время реакции) источника информации, либо уменьшив его внутреннее сопротивление до $\tau_{вт} = I_{ц} \tau_{вт} / (I_{ц} + I_m)$, где $I_{ц}$ — ток, необходимый полезной нагрузке. Первая из указанных мер требует в случае только человеческого управления использования более квали-

фицированного управленческого персонала, а вторая требует увеличения штата сотрудников управления.

Произвольные комбинации источников и нагрузок. Реальные информационные цепи нередко представляют собой сложные переплетения источников и приемников, не сводимые только к последовательным или параллельным соединениям.

Это относится прежде всего к стадным объединениям животных, социальным системам, коллективам, каждый член которых, являясь одновременно приемником и источником информации, обменивается информацией практически с каждым из остальных членов, образуя сложное переплетение информационных связей.

К подобным цепям применимы информационные законы Кирхгофа, первый из которых выражает закон сохранения чувственной информации (принцип непрерывности тока) и формулируется следующим образом: сумма токов, протекающих через любой узел схемы, равна нулю (под узлом понимается любое пересечение или разветвление проводников информации). Второй закон Кирхгофа выражает основное свойство логической информации: суммарные падения напряжения по любым путям между двумя узлами равны между собой (не зависят от пути), т. е. цель не зависит от средств.

Рассмотрим применение этих законов на примере управления деятельностью того или иного подразделения производственного предприятия. С этой целью схематично представим себе, что вся жизнь подразделения может быть условно расчленена на производственную деятельность, идейно-политическую и морально-этическую сферы, а также на сферу охраны труда и быта трудящихся. Предполагая трех ответственных лиц и учитывая взаимную связь и органическое единство перечисленных сфер, можно представить всего два варианта схем управления (рис. 7-6).

На этих схемах индекс 1 присвоен ИДЛ, внутреннему сопротивлению и току административно-хозяйственной сферы, индекс 2 — ИДЛ, внутреннему сопротивлению и току сферы воспитания, индекс 3 — соответствующим характеристикам сферы охраны труда и быта. При этом τ_{12} характеризует информационное сопротивление сферы производства, τ_{13} — информационное сопротивление сферы быта и охраны труда, τ_{23} — информационное сопротивление сферы воспитательной работы.

Схема на рис. 7-6, а соответствует случаю, когда каждый из ответственных лиц непосредственно управляет двумя из трех сфер жизни подразделения: первый — производством и бытовой сферой, второй — производством и сферой воспитания трудящихся, третий — сферой быта и воспитательной работой, причем все они работают в непосредственном контакте друг с другом, так что решение, подготовленное одним из них, приходится согласовывать по меньшей мере с одним из оставшихся, ибо по отношению к каждой нагрузке эти лица соединены попарно последовательно.

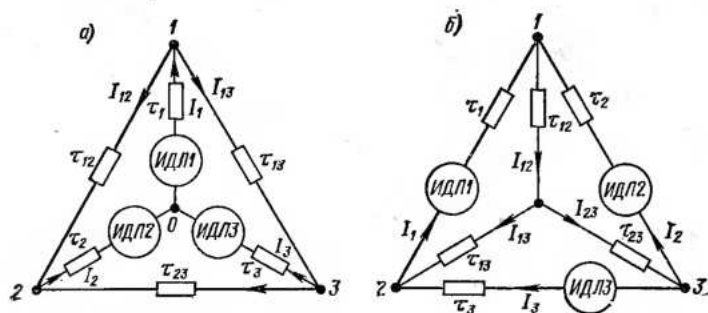


Рис. 7-6. Схемы управления предприятием: а — последовательная; б — параллельная

В отличие от этого схема на рис. 7-6, б допускает независимый (параллельный) выход каждого из них на свои сферы управления по некоторым вопросам, оставляя остальные вопросы для согласованного (последовательного) управления.

Обратимся вначале к первой из этих схем, применив к ней законы Кирхгофа. Первый закон Кирхгофа применительно к узлу 1 разветвления токов (если условиться произвольно обозначенные на рисунке направления токов считать положительными, когда они направлены к узлу, и отрицательными, когда они направлены от узла) дает $I_1 = I_{12} + I_{13}$. Точно так же имеем для узлов 2, 3, 0: $I_2 = I_{12} + I_{23}$; $I_3 = I_{23} - I_{13}$; $I_2 = I_1 + I_3$.

Применяя второй закон Кирхгофа ко всем цепям между узлами 1 и 2 и полагая, что направление ИДЛ источников совпадает с направлениями токов в цепях, получим: $\Delta H_{12} = I_{12}\tau_{12} = h_1 + h_2 - I_1\tau_1 - I_2\tau_2 = h_1 - h_3 - I_1\tau_1 + I_3\tau_3 + I_{23}\tau_{23} = I_{13}\tau_{13} + I_{23}\tau_{23}$.

В принципе эту процедуру можно продолжать для пар узлов: 1 и 3, 2 и 3, 0 и 1, 0 и 2, 0 и 3. Однако в этом нет ни малейшей необходимости, поскольку для определения шести неизвестных токов нам нужно всего шесть уравнений, а мы их уже получили гораздо больше, причем в них вошли все параметры схемы. Решая полученную систему уравнений, можно найти токи во всех ветвях схемы.

Учитывая громоздкость получающихся выражений, мы предоставляем проделать эту операцию тем из читателей, кто интересуется проблемами управления предприятием, а здесь приведем лишь решение упрощенного варианта задачи, когда $h_1 = h_2 = h_3 = h$; $\tau_1 = \tau_2 = \tau_3 = \tau_{\text{вт}}$ и $\tau_{12} = \tau_{13} = \tau_{23} = \tau_{\text{н}}$. В этом случае, как легко сообразить: $I_{13} = 0$; $I_1 = I_3 = I_{12} = I_{23} = 0,5I_2$; $\Delta H_{12} = \Delta H_{32} = I_{12}\tau_{\text{н}} = 2h - (I_1 + I_2)\tau_{\text{вт}} = I_{23}\tau_{\text{н}}$. Отсюда получим: $I_2 = 4h/(\tau_{\text{н}} + 3\tau_{\text{вт}})$; $I_1 = I_3 = I_{12} = I_{23} = 2h/(\tau_{\text{н}} + 3\tau_{\text{вт}})$.

Таким образом, если предположить, что все информационные нагрузки примерно одинаковы, а ответственные лица обладают сопоставимыми ИДЛ и внутренними сопротивлениями, то при работе по схеме рис. 7-6, а возникает реальная угроза, что одна из сфер управления (на схеме τ_{13}) окажется неуправляемой.

Применяя теперь законы Кирхгофа к тем же узлам и ветвям схемы на рис. 7-6, б, получим: $I_{12} = I_1 + I_2$; $I_1 = I_{13} + I_3$; $I_{23} = I_2 + I_3$; $I_{12} = I_{13} + I_{23}$; $\Delta H_{12} = h_1 - I_1\tau_1 = I_{12}\tau_{12} + I_{13}\tau_{13} = I_{12}\tau_{12} + I_{23}\tau_{23} - h_3 + I_3\tau_3 = h_2 - I_2\tau_2 - h_3 + I_3\tau_3$.

Решая эту систему уравнений, можно получить выражения для токов во всех ветвях схемы. По уже оговоренной причине мы не станем этого делать, а вместо этого вновь обратимся к упрощенной при тех же условиях, что и в первом случае, задаче. Отметим сразу, что, несмотря на равенство напряжений источников и одинаковость нагрузок, в схеме рис. 7-6, б никакой симметрии токов в ветвях не наступает, поскольку симметрия любых двух источников всегда нарушается третьим источником. Тем не менее, как легко убедиться, составленная для схемы рис. 7-6, б система уравнений при принятых допущениях совместна только, если $I_{13} = 0$.

В этом случае имеем $I_{12} = I_{23} = 2h/(3\tau_{\text{н}} + \tau_{\text{вт}})$; $I_1 = I_3 = h(\tau_{\text{н}} + \tau_{\text{вт}})/[\tau_{\text{вт}}(3\tau_{\text{н}} + \tau_{\text{вт}})]$; $I_2 = h(\tau_{\text{вт}} - \tau_{\text{н}})/[\tau_{\text{вт}}(3\tau_{\text{н}} + \tau_{\text{вт}})]$.

Сравнивая схемы а и б (рис. 7-6), можно заключить, что обе они при равенстве ИДЛ и внутренних сопротивлений источников и при одинаковых нагрузках приводят

к неуправляемости одной из сфер деятельности подразделения. При этом, если $\tau_n \gg \tau_{вт}$, т. е. если для принятия решения требуется гораздо меньше времени, чем для его исполнения, то схема *а* дает в три раза большие токи информации в нагрузках, чем схема *б*, что обеспечивает ей лучшую управляемость. Зато при этом источники ИДЛ1 и ИДЛ3 в схеме *а* загружены в шесть раз больше, чем в схеме *б*, а источник ИДЛ2 — даже в восемь раз больше.

Если же $\tau_{вт} \gg \tau_n$, т. е. если процедура выработки и согласования управленческих решений требует значительно больше времени, чем их исполнение, то теперь уже схема *б* обеспечивает в три раза большие токи в сферах управления, нежели схема *а*. Соответственно нагрузка источников ИДЛ1 и ИДЛ3 в схеме *б* на треть больше, чем их нагрузка в схеме *а*, а нагрузка источника ИДЛ2 в схеме *б* на треть меньше, чем в схеме *а*, хотя и равна нагрузке источников ИДЛ1 и ИДЛ3 в своей схеме.

В заключение выпишем оба закона Кирхгофа в виде универсальных формул. Итак, для любого узла информационной цепи

$$\sum_{k=1}^n I_k = 0, \quad (7-15a)$$

а для любого замкнутого контура

$$\sum_{k=1}^n \Delta H_k = 0. \quad (7-15б)$$

Смысловая мощность и смысл. Для сопоставления интеллектуальных и творческих возможностей различных людей, организаций и автоматов можно использовать их информационные напряжения, т. е. степень изменения под их управляющим воздействием вероятности того или иного события (достижения цели). Однако такого рода сопоставления продуктивны только при одинаковых информационных токах, генерируемых сравниваемыми объектами.

Действительно, если две проектные организации, обладая одинаковыми потенциальными возможностями (напряжениями), в силу ряда обстоятельств выдают в единицу времени различное количество одинаковой документации, то, очевидно, их трудно признать эквивалентными по творческим возможностям, несмотря на одинаковые информационные напряжения.

Отсюда можно заключить, что интеллектуальная (смысловая) мощность сравниваемых людей или организаций определяется не только их напряжением, но и информационным током, который они способны обеспечить. В силу этого будем считать смысловой мощностью N произведение информационного напряжения и тока:

$$N = \Delta H I, \quad (7-16)$$

где N измеряется в бит²/с.

С помощью (7-16) можно сравнивать по смысловой мощности (по творческим возможностям) людей, управленческие организации и автоматы, отличающиеся как по напряжениям, так и по информационным токам. Так, сравнивая между собой двух изобретателей, первый из которых обладает в два раза более высоким информационным напряжением, чем второй, т. е. его изобретения обеспечивают гораздо больший прирост вероятности нужного исхода управления, чем изобретения второго, но зато эти последние делаются в два раза чаще, чем изобретения первого изобретателя, мы в соответствии с (7-16) вынуждены констатировать одинаковость их смысловых мощностей, несмотря на большую ценность изобретений первого изобретателя. Точно так же мы признаем эквивалентными руководителей разного стиля, если они одновременно достигают цели (например, выполнения плана), управляя одинаковыми со всех точек зрения организациями, хотя один из них отдает лишь весьма редкие, но очень важные по существу указания, а другой склонен вмешиваться в управление по всем незначительным вопросам.

Смысловой мощностью характеризуются не только источники, но и приемники информации (нагрузки). Так, в схеме (см. рис. 7-2) нагрузка согласно (7-16) потребляет с учетом (7-8) смысловую мощность, равную смысловой мощности источника: $N_n = \Delta H I = \tau_n h^2 / (\tau_{вт} + \tau_n)^2$, где

$$I = h / (\tau_{вт} + \tau_n); \quad \Delta H = I \tau_n.$$

В той же схеме на внутреннем сопротивлении источника бесполезно рассеивается смысловая мощность $N_{вт} = \Delta H_{вт} I = h^2 \tau_{вт} / (\tau_{вт} + \tau_n)^2$, тогда как ИДЛ источника располагает смысловой мощностью $N_{идл} = h I = h^2 / (\tau_{вт} + \tau_n)$.

Если, что вполне естественно, трактовать смысловую мощность как вырабатываемый или рассеиваемый в единицу времени смысл S соответствующей деятельности, по-

нимаемой как мера интеллектуального труда, произведенного за время T , то с учетом (7-16) получим (бит²)

$$C = \int_0^T N dt = \int_0^T \Delta H I dt. \quad (7-17)$$

В случае, когда $\Delta H = \text{const}$, имеем с учетом (7-5)

$$C = \Delta H \int_0^T I dt = \Delta H J. \quad (7-18)$$

Определим в качестве примера смысловую мощность и интеллектуальный труд, вложенный машинисткой в печатание того или иного русского текста. Поскольку вероятности (относительные частоты) встретить определенную

Таблица 7-1

Буква или знак	P_k	Буква	P_k	Буква	P_k	Буква	P_k
↔	0,145	р	0,041	я	0,019	х	0,009
о	0,095	в	0,039	ы	0,016	ж	0,008
е, ё	0,074	л	0,036	з	0,015	ю	0,007
а	0,064	к	0,029	ь, ъ	0,015	ш	0,006
и	0,064	м	0,026	б	0,015	ц	0,004
т	0,056	д	0,026	г	0,014	щ	0,003
н	0,056	п	0,024	ч	0,013	э	0,003
с	0,047	у	0,021	й	0,010	ф	0,002

букву, включая промежутки между словами $\langle - \rangle$, в тексте, написанном на русском языке, неодинаковы (см. табл. 7-1), определим потенциал текста как источника буквенной информации, полагая для простоты $P_{\text{усл}} = 1$, т. е. что машинистка работает без ошибок:

$$\Delta H = \sum_{k=1}^{32} P_k \log P_k \approx 4,42 \text{ бит}. \quad (7-19)$$

Если машинистка печатает со скоростью примерно 300 знак/мин, то информационная нагрузка, т. е. время печатания одного знака, составляет $\tau = 0,2$ с. Таким образом, смысловая мощность машинистки составляет

$$N = \Delta H^2 / \tau = 98 \text{ бит}^2/\text{с},$$

а информационный ток

$$I = \Delta H / \tau = 22 \text{ бит}/\text{с}.$$

Значит, проработав непрерывно около 6 ч, она вложит в свою работу смысла, или умственного труда,

$$C = 6 \cdot 3600 N = 2 \cdot 10^6 \text{ бит}^2 \quad (7-20)$$

и переработает информацию

$$J = 6 \cdot 3600 I = 4,76 \cdot 10^5 \text{ бит}.$$

Интересно отметить, что человек читает машинописный текст, по данным [32], в среднем всего лишь в два раза быстрее, чем она, т. е. при чтении букв $\tau = 0,1$ с. Это значит, что при чтении набора машинописных букв (например, бессмысленный текст) человек в среднем развивает смысловую мощность примерно 200 бит²/с, что составляет его физиологический предел, так как он быстрее читать не может. При этом он пропускает информационный ток порядка 50 бит/с, что, по данным [32], составляет физиологический предел человека по пропускной способности зрительного канала.

Научившись определять смысловую мощность, мы можем в наиболее общей форме решить вопрос о коэффициенте полезного действия (КПД) источника информации η . Под КПД мы будем понимать отношение полезной для нагрузки (безызыточной) смысловой мощности N_6 ко всей смысловой мощности источника N

$$\eta = N_6 / N = C_6 / C \quad (7-21)$$

или с учетом (7-17) — отношение соответствующих смыслов за определенное время.

Сравним в качестве примера КПД схем управления a и b (рис. 7-6). Для обеих схем имеем

$$\eta = \frac{I_{12} \Delta H_{12} + I_{13} \Delta H_{13} + I_{23} \Delta H_{23}}{h_1 I_1 + h_2 I_2 + h_3 I_3}, \quad (7-22)$$

что в соответствии с полученными в свое время для этих схем значениями токов дает для схемы a

$$\eta_a = \frac{\tau_n}{\tau_n + 3\tau_{\text{вт}}},$$

а для схемы b

$$\eta_b = \frac{8\tau_{\text{вт}}\tau_n}{(\tau_n + 3\tau_{\text{вт}})(3\tau_n + \tau_{\text{вт}})} = \frac{8\tau_{\text{вт}}\eta_a}{3\tau_n + \tau_{\text{вт}}}.$$

Таким образом, в случае $\tau_n \geq \tau_{nt}$ КПД схемы *a* больше, чем КПД схемы *б*, и близок к единице. В случае $\tau_n = \tau_{nt}$ имеем $\eta_b = \eta_a = 0,5$. Наконец, при $\tau_{nt} \geq \tau_n$ КПД схемы *б* больше, чем КПД схемы *a*, и близок к единице. Интересно, что при $3\tau_n = 7\tau_{nt}$ обе схемы равноценны с точки зрения КПД, но и одинаково малоэффективны, так как $\eta_a = \eta_b = 0,45$.

7-2. ПЕРЕХОДНЫЕ РЕЖИМЫ В ИНФОРМАЦИОННЫХ ЦЕПЯХ

Информационные цепи с памятью. Рассмотрим теперь цепи, обладающие способностью запоминать и хранить информацию. Такой способностью обладают не только живые существа, но и ЭВМ, и автоматические системы управления, и объекты неживой природы. В последнем случае, например, влажная почва «помнит» о прошедшем дожде, а отпечатки на известняке несут информацию о давно исчезнувших животных и растениях.

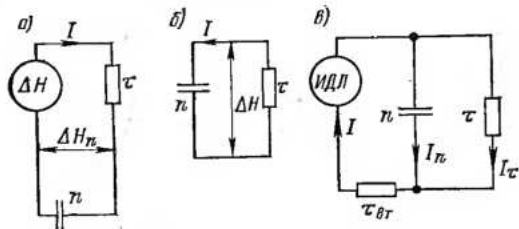


Рис. 7-7. Информационные цепи с памятью:
a и *б* — при последовательной нагрузке;
в — при параллельной нагрузке

Память принято характеризовать емкостью n , которую мы определим здесь как отношение запомненной информации к ее потенциалу (напряжению), т. е. как информацию, необходимую для достижения цели с вероятностью 0,5:

$$n = J/\Delta H. \quad (7-23)$$

Обратимся теперь к информационным цепям, содержащим устройства с памятью. На рис. 7-7, *a* изображена цепь заполнения памяти с емкостью n от источника информации с напряжением ΔH через сопротивление τ , представляющее собой время заполнения одной ячейки памяти. Для непрерывных систем τ соответствует времени заполнения минимально различимой доли памяти.

В такой цепи напряжение источника уравнивается падением напряжения на информационном сопротивлении τ и напряжением на памяти ΔH_n :

$$\Delta H = I\tau + \Delta H_n. \quad (7-24)$$

При этом, поскольку за время заполнения памяти ток будет изменяться, то $J = \int I dt$; тогда из (7-23) следует

$$\Delta H_n = \frac{1}{n} \int I dt. \quad (7-25)$$

Итак, для цепи (рис. 7-7, *a*) имеем уравнение

$$\Delta H = I\tau + \frac{1}{n} \int I dt. \quad (7-26)$$

Решение этого уравнения дает

$$I = \frac{\Delta H}{\tau} \exp\left(-\frac{t}{n\tau}\right). \quad (7-27)$$

Введя понятие постоянной времени $T = n\tau$ заполнения памяти, можем теперь заключить из (7-27), что, если в системе отсутствует специальный регулятор информационного тока, последний весьма велик в начале заполнения памяти (например, в детском возрасте у человека) и уменьшается по экспоненте по мере заполнения, причем тем быстрее, чем меньше T , стремясь к нулю при практически заполненной памяти.

Обычно принято считать, что экспонента достигает установившегося значения практически при $t = (3 \div 5)T$. Следовательно, быстрее всего набираются жизненного опыта люди сообразительные (малое τ), но с небольшой емкостью памяти (малое n). Такие люди раньше других начинают использовать свои ограниченные знания в практических целях. Напротив, очень долго и медленно приобретают практические навыки тугодумы (большое τ) с большой емкостью памяти (большое n), которые обычно становятся эрудитами, часто не приспособленными к самостоятельному использованию своих значительных по объему знаний, поскольку все еще (нередко до конца жизни) пребывают в стадии заполнения своей памяти. Большинство же людей группируется между этими двумя крайними индивидуумами, обладая тем или иным промежуточным значением T . Отметим еще, что сообразительные люди с большой емкостью памяти и тугодумы с малой емкостью памяти созревают в одинаковом возрасте, если, конечно, они

имеют доступ к одинаковым источникам информации, хотя отдача первых неизмеримо больше, чем вторых.

Выражения (7-25) и (7-27) справедливы только в тех случаях, когда информационный ток изменяется по потребности. В практической же деятельности специализированных органов управления и в автоматических системах информационный ток нередко регламентирован и постоянен в процессе заполнения памяти, например при передаче информации по специальным каналам связи, когда ток соответствует пропускной способности канала, или при печатании с постоянной скоростью. В этих случаях вместо (7-25) имеем $\Delta H_n = IT/n$, где T — время заполнения памяти, а вместо (7-27) имеем

$$\Delta H = I(\tau + T/n). \quad (7-28)$$

Выражение (7-28) позволяет определить ток заполнения памяти при заданных времени заполнения и вероятности определенного состояния памяти после заполнения.

Переходя к изображениям по Лапласу и Карсону в (7-25) и (7-27), получим информационную передаточную функцию памяти $I(s)/\Delta H(s) = ns$, являющуюся идеальным дифференцирующим звеном, и передаточную функцию цепи заполнения памяти (рис. 7-7, а)

$$\frac{I(s)}{\Delta H(s)} = \frac{ns}{Ts+1}, \quad (7-29)$$

которые справедливы при любом законе изменения напряжения и тока во времени.

Рассмотрим теперь цепь заполненной памяти (рис. 7-7, б), выдающей информацию на внешнюю нагрузку τ . В этом случае имеем $\Delta H = I\tau = (1/n) \int I dt$,

откуда

$$I = (\Delta H_0/\tau) \exp[-t/(n\tau)], \quad (7-30)$$

что по форме аналогично (7-27), но здесь ΔH_0 — начальное значение напряжения памяти.

Вообще же напряжение памяти уменьшается по мере выдачи информации

$$\Delta H = I\tau = \Delta H_0 \exp\left(-\frac{t}{T}\right) \quad (7-31)$$

и, естественно, обращается в нуль, когда память выдала всю свою информацию по данному поводу и больше не представляет интереса для дела.

В этом случае очевидно, что память считается тем быстрее, чем она меньше и чем быстрее усваивает информацию нагрузка, т. е. чем меньше τ . Отсюда следует, что потенциал, например, учителя в каждый момент времени тем больше потенциала его учеников, чем больше объем его знаний (емкость памяти) и чем тупее его ученики (большое τ). Способные ученики быстро выравнивают свой потенциал с потенциалом учителя, так что разность потенциалов (напряжение) быстро падает в соответствии с (7-31), тем более если объем знаний учителя относительно невелик.

Обратимся теперь к такой цепи (рис. 7-7, в), когда исполнительный орган, помимо выполнения управляющих распоряжений, должен еще их запоминать.

В этом случае

$$h - I\tau_{вт} = I\tau = \frac{1}{n} \int I_n dt \text{ и } I = I_\tau + I_n,$$

откуда

$$\left. \begin{aligned} I_\tau &= \frac{h}{\tau + \tau_{вт}} \left[1 - \exp\left(-\frac{t}{T}\right) \right]; \\ I_n &= \frac{h}{\tau_{вт}} \exp\left(-\frac{t}{T}\right), \end{aligned} \right\} \quad (7-32)$$

где $\tau_{вт}$ — внутреннее сопротивление источника; $T = n\tau\tau_{вт}/(\tau + \tau_{вт})$.

Из (7-32) следует, что в таких условиях ток в нагрузке (рис. 7-8) и напряжение на ней в момент включения источника равны нулю и экспоненциально возрастают до

меры заполнения памяти, достигая наибольшего значения $h/(\tau + \tau_{вт})$ только после ее заполнения. Напротив, ток в памяти максимален ($h/\tau_{вт}$) в момент включения и уменьшается по мере заполнения памяти вплоть до нуля. Что же касается их суммарного тока I , то он все время изменяется от $h/\tau_{вт}$ до $h/(\tau + \tau_{вт})$.

Таким образом, люди, организации и технические системы, которым приходится обучаться в процессе выполнения работы, вначале малоэффективны в работе, так как большую часть информационного тока отправляют в память, но по мере обучения и заполнения памяти они все большую часть управляющего

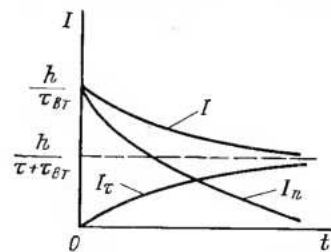


Рис. 7-8. Процессы в цепях с памятью

информационного тока реализуют в деле (т. е. на нагрузке τ). Кроме того, из рис. 7-8 видно, что управление в процессе обучения требует от источника большей мощности, чем управление обучившимся персоналом. При этом, поскольку нагрузка и память работают при одинаковом напряжении, то $I_\tau/I_n = C_\tau/C_n$, а значит, вначале большая часть смысла работы источника направлена на обучение (заполнение памяти) и лишь затем на саму работу, причем этот переход осуществляется тем быстрее, чем меньше информационное сопротивление нагрузки τ и емкость памяти.

Передаточная функция, соответствующая (7-32), имеет вид

$$\frac{I_\tau(s)}{h(s)} = \frac{1}{(\tau + \tau_{\text{вт}})(Ts + 1)},$$

из чего следует, что нагрузка с параллельной памятью является для источника инерционным звеном.

При последовательном соединении двух нагрузок, имеющих память n_1 и n_2 , через них течет одинаковый ток, поэтому

$$\Delta H_1 = \frac{1}{n_1} \int I dt; \Delta H_2 = \frac{1}{n_2} \int I dt,$$

откуда $\Delta H_1/\Delta H_2 = n_1/n_2$, т. е. напряжения на последовательно соединенных нагрузках с памятью обратно пропорциональны их емкостям, а эквивалентная емкость памяти всей схемы

$$\frac{\int I dt}{\Delta H_1 + \Delta H_2} = \frac{n_1 n_2}{n_1 + n_2} = n, \quad (7-33)$$

что для одинаковых нагрузок в два раза меньше, чем емкость каждой из них.

При параллельной работе нагрузок с памятью напряжения на них одинаковы:

$$\Delta H = \frac{1}{n_1} \int I_1 dt = \frac{1}{n_2} \int I_2 dt,$$

откуда заключаем, что в каждый момент времени содержащаяся в нагрузках информация пропорциональна их емкостям, а эквивалентная емкость памяти всей схемы составляет

$$n = \frac{\int I_1 dt + \int I_2 dt}{\Delta H} = n_1 + n_2, \quad (7-34)$$

т. е. емкость памяти параллельно соединенных нагрузок равна сумме их емкостей.

Ригидные информационные цепи. Под ригидностью принято понимать [17] негибкость, неспособность психики человека приспосабливаться к изменяющимся условиям его окружения. Поскольку, однако, эта черта проявляется в управленческих цепях любой природы, включая человеческие коллективы и технические устройства, мы распространим на них термин ригидность для характеристики таких проявлений, как упрямство, привычки, догматизм, консерватизм, неспособность изменять алгоритм своей работы и др. Несмотря на многообразие подобных проявлений, внешне все они выражаются в активном противодействии управлению, т. е. в выработке встречного информационного напряжения, уменьшающего напряжение управления, так что

$$\Delta H_L = L \frac{dI}{dt}, \quad (7-35)$$

где L — ригидность (в с^2), или индуктивность.

Итак, ригидность равна отношению создаваемого ею противонапряжения, которое определяется вызванным ригидностью изменением вероятности достижения цели управления, к скорости изменения тока управления или к ускорению, с которым изменяется информация в каждый момент времени:

$$L = \frac{\Delta H_L}{\frac{dI}{dt}} = \frac{\Delta H_L}{\frac{d^2 J}{dt^2}}.$$

Эта величина связана с временем t_n перехода системы управления на новый порядок (алгоритм) работы. Так, если для перевода вычислительного комплекса ЭВМ на работу по новой программе требуется три рабочих дня по 7 ч в каждом, то его ригидность, как будет показано, имеет порядок $3 \times 7 \times 3600 \tau$, где τ — информационное сопротивление комплекса. Если же эта процедура механизирована и сводится к набору кода соответствующей программы в течение 10 с, то ригидность комплекса составляет уже $L \approx 10\tau \text{ с}^2$.

Подчеркнем, что время перестройки системы на новый алгоритм работы не включает в себя время обучения, которое связано с заполнением памяти ЭВМ или персонала новой информацией. Речь идет о времени, затрачиваемом только на организационно-технические мероприятия и выработку навыков работы при наличии обученного персонала и соответствующих сведений в памяти машины.

Так, для перехода смежных предприятий к работе в производственном объединении не требуется переучивать их персонал; нужно только время для внедрения новых взаимоотношений предприятий в пределах объединения.

На рис. 7-9, а показана схема работы управляющего источника на нагрузку, обладающую ригидностью L и сопротивлением τ . Для этой схемы с учетом (7-35) имеем

$$h - I\tau = L \, dI/dt \text{ или } I = \frac{h}{\tau} [1 - \exp(-t/T)],$$

где $T = L/\tau$ — постоянная времени цепи.

В такой цепи ток в момент подключения управления равен нулю, а затем нарастает по экспоненте вплоть до установившегося значения h/τ , которое не зависит от ригидности. Таким образом, ригидные цепи в установившемся

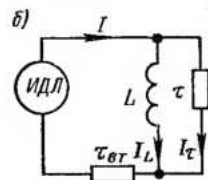
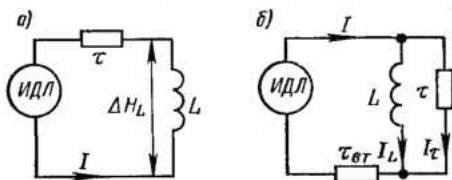


Рис. 7-9. Цепи с последовательной (а) и параллельной (б) ригидностью

режиме (при постоянном алгоритме работы) ничем не отличаются от цепей без ригидности, но входят в этот режим тем дольше, чем больше их ригидность и чем меньше их информационное сопротивление, поскольку переходный режим практически занимает $t_n = (3 \div 5)T$. Отсюда следует также обещанное ранее соотношение $(3 \div 5)L = \tau t_n$.

Лапласово изображение дифференциального уравнения этой цепи дает

$$\frac{I(s)}{h(s)} = \frac{1}{\tau(Ts + 1)},$$

т. е. передаточную функцию инерционного звена.

Отметим, что ток в нагрузке с ригидностью изменяется так же, как ток в нагрузке с параллельной памятью. Эти токи равны при условии равенства информационных сопротивлений τ и постоянных времени T , т. е. при условии $\tau^2 n = L$. Однако при обрыве цепи управления, т. е. при отключении источника, схемы (рис. 7-7, 7-9) ведут себя по-

разному. Если при обрыве цепи управления схема с памятью (рис. 7-7, в) может еще некоторое время управляться по памяти, то цепь с ригидностью (рис. 7-9, а) становится неуправляемой, но зато она прилагает колоссальные усилия для восстановления управления. Действительно, при обрыве ток от источника практически мгновенно падает до нуля, т. е. $(dI/dt) \rightarrow \infty$. Следовательно, согласно (7-35) ригидность создает напряжение $\Delta H \rightarrow \infty$, которое стремится «пробить» разрыв и восстановить связь с источником. Таким образом, ригидные цепи, хотя трудно поддаются перестройке, зато весьма стойки к помехам, вызывающим кратковременные случайные изменения тока управления. Стало быть, проявления ригидности могут быть как положительными, так и отрицательными в зависимости от условий их проявления.

Ригидность отдельного человека проявляется прежде всего в том, что он не сразу воспринимает прогрессивные начинания и согласно (7-35) тем неохотнее, чем большую новизну они вносят, т. е. чем больше dI/dt . Но она же позволяет ему сохранять свои убеждения, несмотря на случайные трудности и злонамеренную подтасовку фактов.

В самонастраивающихся технических системах ригидность проявляется в инерции различной физической природы, которая характеризуется в любых инерционных звеньях постоянной времени T , причем это та же самая постоянная времени, что и в информационной цепи, поскольку речь идет о двух сторонах (информационной и энергетической) одного и того же процесса управления. Это обстоятельство означает, что нормированные (безразмерные) передаточные функции информационной и энергетической цепей одной и той же системы управления одинаковы.

На схеме (рис. 7-9, б) демонстрируется параллельная работа нагрузок, одна из которых практически не обладает памятью и ригидностью, а другая, напротив, обладает столь большой ригидностью, что ее информационное сопротивление и память практически не проявляют себя. Такая схема справедлива, например, когда двум исполнителям поручают одно новое задание, выполнению которого они заранее обучены, но один из них (лабильный) включается в работу немедленно, а другой (ригидный) долго присматривается и настраивается, прежде чем полностью войдет в дело. Зато первого исполнителя работа часто совсем не захватывает и он относится к ней без огонька, а второй в конце

концов настолько увлекается, что работа становится его кровным делом.

Для схемы (рис. 7-9, б): $I = I_\tau + I_L$; $h - I\tau_{\text{вт}} = I_\tau\tau = LdI_L/dt$.

Решения этих уравнений имеют вид

$$\left. \begin{aligned} I_\tau &= \frac{h}{\tau + \tau_{\text{вт}}} \exp\left(-\frac{t}{T}\right); \\ I_L &= \frac{h}{\tau_{\text{вт}}} \left[1 - \exp\left(-\frac{t}{T}\right)\right], \end{aligned} \right\}$$

где $T = L(\tau + \tau_{\text{вт}})/\tau\tau_{\text{вт}}$.

Из полученных решений следует, что в момент включения управляющего источника всю работу принимает на себя лабильный исполнитель (кривая I_τ на рис. 7-10), а ригидный исполнитель практически не дает отдачи (кривая I_L). Затем картина меняется: все большую часть работы берет на себя постепенно освоившийся ригидный исполнитель, а лабильный, пользуясь этим, постепенно разгружает себя и в конце концов сбрасывает всю работу

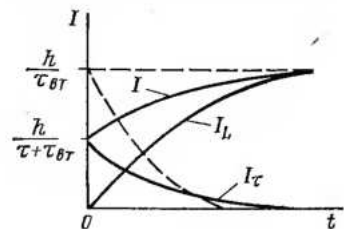


Рис. 7-10. Процессы в ригидных цепях

на своего увлекшегося коллегу. Несмотря на это, общая их производительность все время постепенно нарастает (кривая I на рис. 7-10) от $h/(\tau + \tau_{\text{вт}})$ до $h/\tau_{\text{вт}}$.

Посмотрим теперь, что произойдет в этой системе, если вдруг отключится управление, т. е. иссякнет поток информации. В таком случае $I = 0$; $-I_L = I_\tau$; $I_\tau\tau = LdI_L/dt$, откуда получим

$$I = \frac{h}{\tau_{\text{вт}}} \exp(-t/T_1),$$

где $T_1 = L/\tau$.

Это значит, что при отключении управления в установившемся режиме ригидный исполнитель не только некоторое время будет продолжать работу, лишь постепенно экспоненциально снижая активность (штриховая кривая на рис. 7-10), но заставит работать в таком же темпе и своего лабильного коллегу, который перед этим совсем было перестал работать.

При последовательном соединении нагрузок с ригидностью L_1 и L_2 через них течет общий ток I , поэтому

$$\Delta H_1 = L_1 dI/dt; \Delta H_2 = L_2 dI/dt,$$

откуда $\Delta H_1/\Delta H_2 = L_1/L_2$, т. е. их напряжения пропорциональны ригидностям, а эквивалентная ригидность схемы равна сумме ригидностей отдельных нагрузок

$$L = \frac{\Delta H_1 + \Delta H_2}{dI/dt} = L_1 + L_2. \quad (7-36)$$

При параллельной работе ригидные нагрузки имеют одинаковое напряжение

$$\Delta H = L_1 dI_1/dt = L_2 dI_2/dt,$$

откуда следует, что их эквивалентная ригидность составляет

$$L = \frac{\Delta H}{d(I_1 + I_2)/dt} = \frac{L_1 L_2}{L_1 + L_2}. \quad (7-37)$$

Таким образом, при выдаче нового задания одновременно двум одинаковым организациям их суммарная ригидность в два раза меньше, чем у каждой организации в отдельности.

До сих пор речь шла о собственной ригидности объекта управления, являющейся отражением его внутренних особенностей. Между тем можно обнаружить проявления ригидности, вызванные внешними влияниями, когда приверженность определенному поведению (алгоритму) диктуется не привычками и собственными убеждениями, а тем, что «все так делают», или тем, что так поступает лицо (организация), пользующееся непререкаемым авторитетом у объекта управления. Такого рода взаимовлияние информационных цепей мы будем называть взаимной ригидностью. Влияние взаимной ригидности может проявляться не только в согласовании алгоритмов работы различных цепей, но и противоположным образом, когда одна из нагрузок умышленно пользуется алгоритмом, отличным от алгоритмов работы нагрузок в других аналогичных цепях, по конкурентным соображениям, из духа противоречия и антипатии, либо из стремления к оригинальности.

Посмотрим, как будет вести себя нагрузка, связанная посредством взаимной ригидности $L_{\text{вз}}$ с информационным током I_2 в смежной цепи (рис. 7-11).

Если цепи склонны к взаимному согласованию алгоритмов работы, то будем брать ригидность $L_{вз}$ со знаком минус; если же они пребывают в конфликтных отношениях и мешают друг другу, то будем брать $L_{вз}$ со знаком плюс, так что наводимое током I_2 напряжение в цепи тока I_1 определится выражением

$$\Delta H_{вз} = L_{вз} dI_2/dt. \quad (7-38)$$

Итак, с учетом (7-38) имеем для схемы рис. 7-11

$$h = I_1\tau + L_{вз} \frac{dI_2}{dt}; \quad I_1 = \frac{h - L_{вз} \frac{dI_2}{dt}}{\tau}. \quad (7-39)$$

Из (7-39) следует, что в установившемся режиме ток I_1 определяется только информационным сопротивлением своей собственной цепи и не зависит от I_2 . Если же I_2 изменяется с постоянной скоростью $dI_2/dt = \xi = \text{const}$, т. е. если цепь I_2 постоянно эволюционирует, то согласно (7-39) ток I_1 больше или меньше установившегося на $L_{вз}\xi/\tau$ в зависимости от того, согласно или встречно текут токи I_1 и I_2 . При

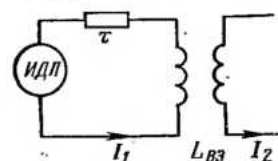


Рис. 7-11. Взаимная ригидность цепей

согласной работе цепей ($L_{вз} < 0$) увеличение тока I_2 приводит к появлению в цепи тока I_1 дополнительного напряжения $\Delta H_{вз}$, которое прибавляется к ИДЛ источника. Это значит, что при работе, например, двух организаций по схожим алгоритмам успех одной из них стимулирует деятельность другой, так как увеличивает вероятность достижения цели. Напротив, трудности в работе одной из них (уменьшение I_2) уменьшают вероятность достижения цели другой организацией и снижают потенциал управления. Зато если эти организации разрабатывают альтернативные варианты ($L_{вз} > 0$), то, наоборот, неуспех одной из них повышает шансы другой добиться успеха.

Информационные цепи с памятью и ригидностью. В самом общем случае информационные цепи обладают и сопротивлением, и памятью, и ригидностью, поэтому представляет интерес рассмотреть работу такой цепи (рис. 7-12).

На рис. 7-12, а изображена схема заполнения памяти ригидной нагрузки, что соответствует процедуре заучивания непривычного материала, например иностранных слов.

Уравнение такой цепи

$$h = I\tau + L \frac{dI}{dt} + \frac{1}{n} \int I dt$$

имеет двоякое решение:

$$I = \frac{h\delta}{\tau\sqrt{\delta^2-1}} \left[\exp \frac{(-\delta + \sqrt{\delta^2-1})t}{T} - \exp \frac{(-\delta - \sqrt{\delta^2-1})t}{T} \right], \quad (7-40a)$$

если $\delta > 1$, где $\delta = \frac{\tau}{2} \sqrt{n/L}$, $T = \sqrt{nL}$, или

$$I = \frac{2h\delta}{\tau\sqrt{1-\delta^2}} \exp \left(-\frac{\delta t}{T} \right) \sin \left(\frac{t\sqrt{1-\delta^2}}{T} \right), \quad (7-40b)$$

если $\delta < 1$.

Первое решение соответствует сравнительно малой ригидности $4L < \tau^2 n$, ограничивающей скорость усвоения

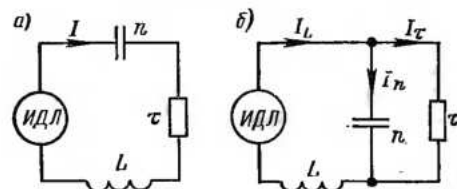


Рис. 7-12. Ригидные цепи с последовательной (а) и параллельной (б) памятью

материала лишь в начале процесса (кривая $\delta > 1$ на рис. 7-13, а), а затем (после приспособления к работе) практически не влияющей на производительность запоминания, которая в этом случае определяется лишь остатком свободного места в памяти, причем в этих условиях однажды заученный материал не забывается.

Второе решение соответствует большой ригидности $4L > \tau^2 n$ (кривая $\delta < 1$, рис. 7-13, а), которая вначале препятствует усвоению материала, а затем, когда эта работа становится привычной, продолжает ее сверх необходимого, что приводит к переполнению памяти, например, иностранными словами, знание которых необязательно для чтения русских текстов.

Естественно, через некоторое время память освобождается от ненужного багажа, чему ригидность вначале препятствует, а потом усугубляет этот процесс сверх меры, так что забывается и часть нужных слов, что требует нового запоминания и т. д. В результате этого возникает

колебательный процесс (рис. 7-13, а кривая $\delta < 1$) за-
поминания с частотой $\sqrt{1-\delta^2}/T$ и затуханием δ/T , когда
периоды активного усвоения материала сменяются перио-
дами частичного забывания. Этот процесс постепенно за-
тухает, и память в конце концов заполняется.

Анализ полученных выражений показывает, что плав-
ный процесс заполнения памяти (заполнение с одного раза)
имеет место у неконсервативных людей и организаций
(малое L) с большой емкостью памяти и замедленным вос-
приятием информации (большое τ). Напротив, колебатель-
ный процесс заполнения памяти свойствен относительно

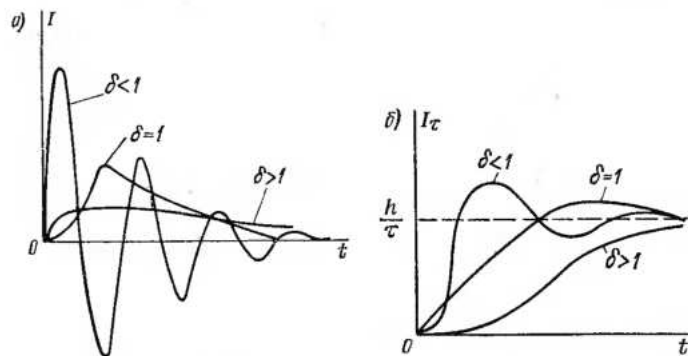


Рис. 7-13. Процессы в ригидных цепях с последовательной (а)
и параллельной (б) памятью

консервативным людям и организациям (большое L) с о-
граниченной емкостью памяти при быстром восприятии
информации (малое τ). Пожалуй, плавный процесс запол-
нения памяти характерен в основном для многолюдных
организаций, которые обладают большим объемом памяти
и медленнее усваивают информацию, чем отдельный че-
ловек, которому, скорее, свойствен колебательный про-
цесс ввиду быстрого восприятия и меньшей емкости памяти.

Тем не менее самое быстрое заполнение памяти имеет
место при $\delta = 1$, т. е. $4L = \tau^2 n$, когда из (7-40б) предель-
ным переходом получается

$$I = \frac{2h\delta t}{\tau T} \exp\left(-\frac{t}{T}\right). \quad (7-40в)$$

В этом случае возникает кратковременный мощный
всплеск тока (кривая $\delta = 1$, рис. 7-13, а), быстро запол-
няющий память.

Рассмотрим работу управляющего источника на ри-
гидную нагрузку с параллельной памятью (рис. 7-12, б).
В этом случае память существует для сглаживания неравно-
мерного поступления информации от источника и работает
параллельно с исполнительными органами. Подобная схема
обычна для большинства систем управления. Уравнения,
характеризующие ее, имеют вид:

$$I_L = I_n + I_\tau;$$

$$h = L dI_L/dt + I_\tau \tau; \quad I_\tau \tau = \frac{1}{n} \int I_n dt.$$

Решение при $T = \sqrt{nL}$ и $\delta = \frac{1}{2\tau} \sqrt{L/n}$ имеет либо форму

$$I_\tau = \frac{h}{\tau} \left[1 - \frac{\delta + \sqrt{\delta^2 - 1}}{2\sqrt{\delta^2 - 1}} \exp \frac{t(-\delta + \sqrt{\delta^2 - 1})}{T} + \right. \\ \left. + \frac{\delta - \sqrt{\delta^2 - 1}}{2\sqrt{\delta^2 - 1}} \exp \frac{t(-\delta - \sqrt{\delta^2 - 1})}{T} \right] \quad (7-41а)$$

при $\delta > 1$, либо форму

$$I_\tau = \frac{h}{\tau} \left[1 + \left(\frac{\delta}{\sqrt{1-\delta^2}} \sin \frac{t\sqrt{1-\delta^2}}{T} - \cos \frac{t\sqrt{1-\delta^2}}{T} \right) \exp\left(-\frac{\delta t}{T}\right) \right] \quad (7-41б)$$

при $\delta < 1$.

Из полученных решений следует (рис. 7-13, б), что при
малых нагрузке и памяти и большой ригидности ($L > 4\tau^2 n$)
система управления достигает своей проектной произво-
дительности труда $I_\tau = I_L = h/\tau$ согласно (7-41а) в ре-
зультате плавного и постепенного наращивания темпов
труда по мере освоения нового алгоритма работы (кривая
 $\delta > 1$). Для относительно больших нагрузки и памяти
при ограниченной ригидности ($L < 4\tau^2 n$) из (7-41б) следует,
что систему будет некоторое время лихорадить (кривая
 $\delta < 1$), прежде чем она достигнет проектной производи-
тельности труда, причем частота колебаний производи-
тельности составит $\sqrt{1-\delta^2}/T$, а затухание δ/T . Интересно
отметить, что в случае $\delta = 1$, т. е. когда $L = 4\tau^2 n$, имеет
место самое быстрое достижение установленной произво-
дительности труда (кривая $\delta = 1$), которое описывается
предельным переходом в (7-41б):

$$I_\tau = \frac{h}{\tau} \left[1 + \left(\frac{t}{T} - 1 \right) \exp\left(-\frac{t}{T}\right) \right]. \quad (7-41в)$$

При этом (7-41в) дает лишь однократный небольшой (примерно 14%-ный) всплеск производительности труда с последующим постепенным уменьшением ее до нормы, что является, несомненно, оптимальным вариантом среди всех вариантов переходного режима.

Отметим еще раз, что обе рассмотренные здесь схемы (рис. 7-12) характерны для процессов обучения новому. Однако схема *а* соответствует заучиванию (записи в памяти) нового алгоритма работы, вроде освоения студентами лекционного материала без его практического использования, а схема *б* — освоению нового в процессе его использования, т. е., если обратиться к примеру со студентами, — освоению материала с учетом упражнений и лабораторных занятий. Применительно к процессу обучения студентов таким образом время затухания информационного тока на рис. 7-13, *а* соответствует времени освоения только лекционного теоретического материала, а время установления тока на рис. 7-13, *б* — всему сроку обучения студентов в вузе.

Выражения (7-40) и (7-41) описывают значения токов (информационных производительностей) в любой будущий момент времени относительно момента $t = 0$. Значит, решая исходные дифференциальные уравнения схем, мы тем самым прогнозируем их будущие состояния по заданным начальным условиям. Следовательно, алгоритм решения дифференциального уравнения есть алгоритм прогностического суждения, которое применительно, например, к процессу обучения позволяет по заданному содержанию ИДЛ обучения и известным параметрам τ , L , n обучаемого предсказать ход этого процесса. Для систем, описываемых линейными дифференциальными уравнениями, наиболее удобной формой записи оператора прогностического суждения являются передаточные функции систем. Так, для схемы рис. 7-12, *а* передаточная функция имеет вид

$$\frac{I(s)}{h(s)} = \frac{ns}{1 + \tau ns + nLs^2}, \quad (7-40r)$$

а для схемы рис. 7-12, *б*

$$\frac{I(s)}{h(s)} = \frac{1}{\tau \left(1 + \frac{Ls}{\tau} + nLs^2 \right)}, \quad (7-41r)$$

где $h(s)$ и $I(s)$ — изображения ИДЛ по Лапласу или Карсону; s — комплексный оператор дифференцирования.

Однако передаточные функции оперируют не самими процессами, а их изображениями в комплексной плоскости. По этой причине прогнозированию с их помощью всегда сопутствуют прямое и обратное преобразования Лапласа, что, как мы увидим ниже, связано еще с процедурой преобразования информации (усиления смысла).

Из (7-40) и (7-41) следует, что процесс обучения как будто не имеет конца, т. е. соответствующий ток устанавливается только при $t = \infty$, что вполне соответствует слову «век живи — век учись». Однако практически чем больше состояний изучаемой системы человек запомнил и усвоил, тем меньше вероятность встретить незнакомую ситуацию, так что в конце концов этой вероятностью можно пренебречь. Таким образом, обучение может быть прекращено при $t = (3 \div 5) T$, где $T = \sqrt{L/n}$, если δ не слишком сильно отличается от единицы. В противном случае, если δ значительно больше единицы, обучение может оказаться слишком долгим, чтобы иметь смысл, а если δ много меньше единицы, обучение может оказаться неуспешным ввиду чрезмерной возбужденности обучаемого и невозможности длительного сосредоточения его внимания на определенном вопросе.

Может показаться, что по завершении обучения информационный ток определяется только сопротивлением цепи и не зависит ни от памяти, ни от ригидности. Это действительно было бы так, если бы человек обучался только поведению в одной и той же неизменной ситуации, например учился нажимать одну и ту же кнопку при загорании одной и той же лампочки. Однако обучение даже рутинному труду предусматривает реакцию человека на более или менее значительный набор различных ситуаций, которые могут периодически повторяться. Если все они повторяются с одинаковой частотой ω_0 , то применительно к схеме обучения (рис. 7-12, *б*) $h = h_0 \sin \omega_0 t$, а $I_\tau = h_0 A(\omega_0) \sin[\omega_0 t + \theta(\omega_0)]$, где $A(\omega_0) = \frac{1}{\tau \sqrt{(1 - nL\omega_0^2)^2 + (L\omega_0/\tau)^2}}$ — модуль передаточной функции (7-41r) при $s = j\omega_0$; $\theta(\omega_0) = \frac{L\omega_0}{\tau(nL\omega_0^2 - 1)}$ — ее фаза.

Нередко, однако, изучавшиеся ситуации повторяются с разными частотами; тогда $h = \sum_{k=1}^m A_{k0} \sin(\omega_k t + \theta_{k0})$, а

$I_\tau = \sum_{k=1}^m A_k(\omega_k) \sin[\omega_k t + \theta_k(\omega_k)]$, где m — число ситуаций с различными частотами повторения.

Наконец, если возникают единичные не изучавшиеся ранее ситуации, то процесс обучения протекает непрерывно и описывается посредством передаточной функции (7-41г) при $s = j\omega$, где ω непрерывно изменяется от $-\infty$ до $+\infty$:

$$I_\tau(s) = \frac{h(s)}{\tau(1 + Ls/\tau + nLs^2)}.$$

Преобразование информации. До сих пор мы рассматривали только цепи, в которых информация не подвергалась никаким преобразованиям, т. е. в источнике и нагрузке протекал один и тот же информационный ток. Точнее сказать, в таких цепях источник и приемник имели дело с одними и теми же носителями информации. Между тем весьма распространены иерархические структуры управления, в которых на различных уровнях, осуществляющих, по существу, одни и те же функции, используются различные носители информации. При этом можно выделить два типа преобразований: трансформация и усиление смысла информации.

Под усилением понимается изменение первоначального смысла информации (увеличение или уменьшение), которое имеет место при управлении источниками информации, например, при тиражировании информации, при переводе с языка на язык, кодировании и т. д.

Трансформация представляет собой частный случай усиления с коэффициентом усиления смысла, равным единице, т. е. это такое преобразование информации, в процессе которого сохраняется ее первоначальный смысл, хотя она и представлена другими носителями. Это всегда имеет место при переходе от более мелких единиц измерения к более крупным или наоборот (например, при переходе от исчисления массы груза в тоннах к исчислению его в килограммах или — объема текста в знаках к исчислению его в печатных листах и т. п.). При этом, поскольку смысл C_2 трансформированной информации равен первоначальному ее смыслу C_1 , то с учетом (7-18) имеем

$$Y_1/Y_2 = H_2/H_1, \quad (7-42)$$

т. е. при трансформации усиление (ослабление) информации в точности равно ослаблению (усилению) ее сути.

В схемах преобразования информации (рис. 7-14) цепи источника ИДЛ h_1 и приемника τ информации разделены усилителем, который изображается квадратом со знаком неравенства, указывающим соотношение между смысловыми мощностями приемника и источника.

Так, в схеме *a* смысловая мощность N_1 источника меньше смысловой мощности N_2 приемника, а на схеме *б* наоборот.

Поскольку усилитель всегда располагает некоторым собственным запасом смысла N_0 (рис. 7-14, *a*), то его источник h_0 обычно на схемах не изображается (рис. 7-14, *б*).

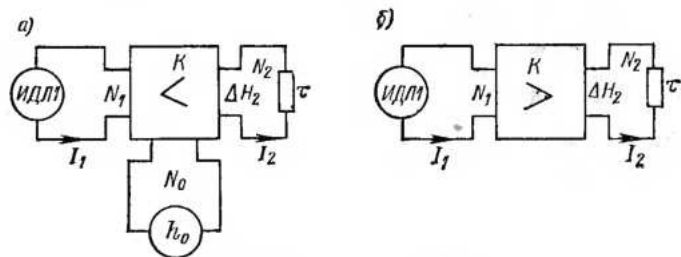


Рис. 7-14. Схемы преобразования информации

Схема *a*, в которой коэффициент усиления $K > 1$, соответствует, например, разговору двух людей через переводчика, когда крайне важная информация источника непосредственно для приемника имеет мало смысла, т. е. источник обладает для него малой смысловой мощностью N_1 . В таких условиях владеющий алгоритмом перевода со смысловой мощностью N_0 переводчик выполняет функцию усилителя смысловой мощности источника до уровня N_2 , причем $N_2 = N_1 + N_0$, поскольку после перевода информация приобретает для приемника смысловую мощность N_2 . Усилитель смысловой мощности (переводчик) характеризуется в этом случае коэффициентом усиления смысловой мощности (смысла):

$$K_N = N_2/N_1 = 1 + N_0/N_1. \quad (7-43)$$

Отметим, что при переводе может быть и $K_N < 1$, если информация сама по себе малозначительна.

То же самое происходит при декодировании, решении уравнений и т. п., когда закодированная или заданная уравнением информация для непосвященного имеет мало смысла и приобретает его только после приведения туда смысла алгоритма декодирования или решения. При этом

усиление смысла происходит только для непосвященного в тайны шифра потребителя информации; вообще же здесь действует закон сохранения смысла, согласно которому смысл расшифрованной информации равен смыслу нерасшифрованной информации плюс смысл алгоритма расшифровки. Отметим, что с излагаемых позиций ЭВМ есть не что иное, как усилитель смысла вводимой информации за счет прибавления к нему (или вычитания) смысла алгоритма своей работы.

Рассмотрим такую ситуацию. Предположим, что заброшенный на вражескую территорию разведчик должен по радио сообщить, какой из десяти ранее оговоренных пунктов он выбрал для своего пребывания. Поскольку возможен радиоперехват сообщения противником, разведчик пользуется шифром (табл. 7-2), где порядковому номеру каждого из 10 пунктов соответствует комбинация двух из пяти букв русского алфавита.

Таблица 7-2

Пункт	0	1	2	3	4	5	6	7	8	9
А	Х	Х	Х	Х	—	—	—	—	—	—
Б	Х	—	—	—	Х	Х	Х	—	—	—
В	—	Х	—	—	Х	—	—	Х	Х	—
Г	—	—	Х	—	—	Х	—	Х	—	Х
Д	—	—	—	Х	—	—	Х	—	Х	Х

Так, пункту № 0 соответствует комбинация АБ, пункту № 1 — комбинация АВ и т. д. В результате в эфир передается непонятное противнику сочетание букв вместо номера пункта. Каждая буква такого сочетания несет в себе, как показывалось ранее (7-19), 4,42 бит потенциала и 1 бит информации, а все сочетание содержит для непосвященного 2 бита информации и $C_1 = 8,8$ бит² смысла. После декодирования вместо каждой пары букв адресат получает одну цифру, которая несет потенциал $\log 10 = 3,32$ бит и 1 бит информации, т. е. имеет смысл $C_2 = \log 10 = 3,32$ бит.²

Таким образом, коэффициент усиления шифровальщика составит $K_c = c_2/c_1 = 0,376$, т. е. роль шифровальщика состоит в уменьшении смысла информации для ее восприятия адресатом. Соответственно сам разведчик, шифруя сообщение, завышает его смысл, чтобы затруднить

восприятие информации противником. При расшифровке убывает потенциал сообщения, как и информация, так что коэффициент усиления по напряжению составляет $K_H = 3,32/4,42 = 0,75$, а коэффициент усиления информации $K_I = 1/2$ при общем соотношении

$$K_c = K_I K_H. \quad (7-44)$$

Для сравнения представим себе, что в тыл противника заброшены два разведчика, которые должны разместиться в пунктах А — Д, причем все возможные расположения разведчиков пронумерованы в табл. 7-2 цифрами от 0 до 9. В этом случае в эфир посылается только одна цифра, означающая конкретное расположение разведчиков. Для несведующего человека она несет в себе информацию $J_1 = 1$ бит и потенциал $H_1 = \log 10$, а для адресата после расшифровки выбор из десяти сочетаний пунктов несет $J_2 = 2$ бит, $H_2 = \log 10$. Соответственно $C_1 = \log 10$, $C_2 = 2 \log 10$.

Таким образом, $K_c = 2$; $K_H = 1$; $K_I = 2$, т. е. этот шифр легче распознается, чем в случае заброски одного разведчика.

В рассмотренных случаях усилению (ослаблению) подвергались как напряжение, так и информация. Между тем нередки случаи, когда усиление смысловой мощности происходит при постоянстве напряжения или информационного тока. Например, при розыске человека, когда имеется его словесный портрет, нередко приглашают художника, который по этому описанию рисует портрет. Естественно, что при такой операции оба портрета (словесный и живописный) в одинаковой степени соответствуют оригиналу, т. е. имеют одинаковое напряжение как источники информации, а роль художника как усилителя смысловой мощности сводится лишь к увеличению информационного тока. В этом случае художник представляет собой усилитель, так как $K_H = 1$, а $K_N = K_I$ причем $K_I > 1$, поскольку живописный портрет лаконичнее словесного. В результате словесное описание читается медленно, а живописный портрет воспринимается почти мгновенно, так что если сравнить их мгновенные информационные токи, то при $\tau_2 < \tau_1$: $(I_2 = H/\tau_2) > (I_1 = H/\tau_1)$.

Напротив, если за счет двукратного уменьшения быстрогодействия, например, измерительного прибора достигается такое увеличение точности измерений, что новая шкала содержит в тех же пределах число делений, равное квад-

рату числа делений прежней шкалы, то потенциал прибора возрастает в два раза, а информационный ток остается прежним из-за двукратного увеличения времени измерений. Таким образом, смысловая мощность возрастает только за счет $K_N = K_H = 2$ при $K_I = 1$.

В заключение отметим весьма важное обстоятельство: усилитель смысловой мощности всегда является направленным звеном, так что изменение нагрузки в выходной цепи никак не влияет на параметры входной цепи. Действительно, сколько бы людей ни слушало выступающего в качестве усилителя смысла переводчика (т. е. каким бы ни было информационное сопротивление его нагрузки), это не имеет никакого значения для источника иностранной речи, поскольку он имеет дело только с переводчиком как с потребителем своей информации вне зависимости от того, усиливает ли он ее или нет.

7.3. МЕТОДОЛОГИЯ СИСТЕМНОГО АНАЛИЗА

Весь предыдущий материал гл. 7 можно истолковать как иллюстрацию применения диалектической логики, математизированной на информационной основе, к тем или иным конкретным системам. Ниже излагается аппарат этой логики как универсальное орудие исследования любых систем и задач управления.

Классическая, формальная логика в основе своей бинарна, т. е. двузначна, поскольку признает либо абсолютную истину, либо абсолютную ложь. Между тем в реальных задачах практической деятельности мы никогда не располагаем абсолютно достоверной информацией, т. е. имеем дело с относительной истиной, достоверность которой колеблется в общем случае от нуля до 100%, никогда не достигая этих пределов. Это значит, что ни классическая, ни следующая ей математическая логика не пригодны для решения задач управления во всем объеме и многообразии и что возникла потребность в бесконечнозначной логике, имеющей дело с эволюционирующей относительной истиной. Действительно, о какой эффективности управления может идти речь, если классическая логика приводит наше мышление к множеству парадоксов, вроде парадокса лжеца: «То, что я говорю, — ложь?» Здесь, если иметь в виду лишь два состояния (абсолютная истина — абсолютная ложь), то, сказав правду, я тем самым солгал, а сказав ложь, произнес истину. Зато, если пользоваться

бесконечнозначной относительной истиной, то фраза в кавычках должна означать: «То, что я говорю, — в какой-то мере ложь» — и ситуация перестает быть парадоксальной. В этом случае, сказав правду на 70%, я тем самым солгал на 30% и никакого противоречия не возникает. Аналогичным образом разрешаются и все другие известные парадоксы.

Приступая к построению бесконечнозначной непрерывной логики [17], следует отдавать себе отчет в том, что для нее не будет справедлив (по меньшей мере в традиционной форме) закон исключенного третьего, поскольку он относится к выбору между истиной и ложью, а в непрерывной логике приходится выбирать какие-то состояния из бесконечного множества состояний относительной истины.

Вместе с тем закон обратной пропорциональности объема понятия и его сущности в основе своей непрерывен и его действие должно распространиться на непрерывную логику. Запишем этот закон в форме

$$H_n = J/n, \quad (7-45)$$

где H_n — сущность понятия; n — объем понятия, т. е. число знаков, необходимое для обозначения всех объектов, охватываемых понятием; J — число аналогичных знаков (слова, фразы), передающих понятие.

Поскольку знаки являются объектом чувственного восприятия, то в дальнейшем будем называть J чувственной информацией, или просто информацией и измерять ее в битах. Что касается объема понятия, то, поскольку понятие формируется на основе сведений, содержащихся в нашей памяти, n есть, по существу, объем (емкость) памяти, занятый этими сведениями.

Согласно (7-45), например, сущность понятия «цвет» равна $1/7$, поскольку, если ограничиться цветами спектра, $n = 7$ и $J = 1$, т. е. одному слову, выражающему понятие, соответствует семь слов, выражающих объем понятия (названия цветов спектра).

Сущность понятия H_n в отличие от J является логической информацией, т. е. не может быть объектом непосредственного чувственного восприятия, но, как и всякая информация, оно должно выражаться в битах.

Возьмем (7-45) за основу и станем конструировать логику управления как методологию, т. е. как диалектиче-

скую логику, путем последовательного выполнения требований законов диалектики применительно к суждению.

Первым в числе этих законов назовем закон всеобщей взаимосвязи и взаимозависимости явлений. Нетрудно подметить, что (7-45) не удовлетворяет этому закону, рассматривая лишь собственную суть понятия вне системы взаимодействий со смежными понятиями. Чтобы учесть требования этого закона, необходимо переписать (7-45) в иной форме, сохранив принцип обратной пропорциональности:

$$H_{n1} = J_2/n_{11} + J_2/n_{12} + J_3/n_{13} + \dots, \quad (7-45a)$$

где системная суть H_{n1} понятия складывается из его собственной сути (7-45) и взаимных сущностей, образованных по тому же принципу на основе взаимодействия с другими понятиями, входящими в ту же систему. Первый индекс при символах означает номер понятия в этой системе, а второй индекс — номер понятия, с которым взаимодействует данное, так что n_{ik} означает взаимный объем i -го и k -го понятий. Это суждение соответствует обычному правилу формальной логики, по которому в любом определении H_1 должны присутствовать, во-первых, его родовая принадлежность J_1 и, во-вторых, видовые отличия от J_2, J_3 и т. д.

Перейдем теперь к требованиям, вытекающим из закона изменчивости «все течет, все изменяется». Понятно, что (7-45a) не удовлетворяет этому закону, поскольку ничего не говорит о процессах становления понятий. Чтобы описать эти процессы, нам придется ввести символ (оператор) изменения d/dt , где числитель означает изменение само по себе, а знаменатель указывает, что это изменение происходит во времени t , так что, например, dJ/dt означает изменение информации во времени. При этом, если происходит накопление информации, то изменение будем считать положительным, а если происходит убыль информации, то — отрицательным. Отметим, что поскольку всякое изменение отрицает стабильность, т. е. ранее неизменное, то закон изменчивости с равным основанием мог бы быть назван законом отрицания, причем символ d/dt есть синоним отрицания, выражаемого в речи частицей «не».

С учетом введенных обозначений сущность H_τ процесса изменения (эволюции) понятия должна быть обратно пропорциональна уже не объему понятия, а изменению этого

объема $dn/dt = 1/\tau$, так что

$$H_\tau = \tau dJ/dt. \quad (7-46)$$

В дальнейшем τ будем именовать информационным сопротивлением понятия.

Выражение (7-46) так же, как и (7-45), не удовлетворяет закону всеобщей взаимосвязи и взаимозависимости явлений, поэтому его следует преобразовать к виду, аналогичному (7-45a):

$$H_{\tau 1} = \tau_{11} dJ_1/dt + \tau_{12} dJ_2/dt + \tau_{13} dJ_3/dt + \dots, \quad (7-46a)$$

где системная суть $H_{\tau 1}$ эволюции понятия складывается из собственной сути эволюции (7-46) и взаимных сущностей эволюций смежных понятий по тому же принципу, что и в (7-45a). При этом τ_{kk} означает собственное информационное сопротивление, а τ_{ik} — взаимное информационное сопротивление i -го и k -го понятий.

Далее, учтем требования, вытекающие из закона отрицания отрицания. Поскольку уже отмечалось, что закон изменчивости является по своей сути законом отрицания, то закон отрицания отрицания следует трактовать как двукратное применение закона изменчивости, т. е. как применение закона изменчивости к самому закону изменчивости. С формально-символической точки зрения это означает применение оператора d/dt к самому оператору d/dt , т. е. $d(d/dt)/dt$, или d^2/dt^2 , так что, например, d^2J/dt^2 означает изменение эволюции информации во времени. Таким образом, теперь сущность H_L изменения эволюции понятия должна быть обратно пропорциональна изменению эволюции объема понятия $d^2n/dt^2 = 1/L$, так что

$$H_L = L d^2J/dt^2, \quad (7-47)$$

где L будем именовать ригидностью (негибкостью) понятия.

Приспосабливая (7-47) к требованиям закона всеобщей взаимосвязи и взаимозависимости явлений, необходимо переписать его в форме, аналогичной (7-45a) и (7-46a):

$$H_{L1} = L_{11} d^2J_1/dt^2 + L_{12} d^2J_2/dt^2 + L_{13} d^2J_3/dt^2 + \dots, \quad (7-47a)$$

где системная суть H_{L1} изменения эволюции понятия складывается из собственной сути изменения эволюции (7-47) этого понятия и взаимных сущностей изменения эволюций смежных понятий. При этом L_{kk} означает собственную ригидность данного понятия, а L_{ik} — взаимную ригидность i -го и k -го понятий.

Закон отрицания мог бы быть применен к закону отрицания отрицания, что привело бы к закону отрицания отрицания отрицания, символизируемому d^3/dt^3 , и т. д. Однако диалектика не напрасно не содержит законов тройного и более высоких отрицаний, поскольку это было бы слишком громоздко и обременительно для мышления. Наше мышление действует экономно, и все явления, выходящие за рамки закона отрицания отрицания, разбивает на транзитивные формы: $H_{L1} = L_1 d^2 J_1 / dt^2$; $H_{L2} = L_2 d^2 H_{L1} / dt^2$ и т. д., которые в совокупности передают сколь угодно сложные процессы.

Следующий закон диалектики, требования которого необходимо учесть, — это закон единства противоположностей. Противоположностями в нашем случае являются момент относительной стабильности понятия (7-45а) — тезис и момент относительной изменчивости (7-46а) — антитезис, опосредствованные отрицанием отрицания (7-47а), которые в соответствии с применяемым законом диалектики следует рассматривать в неразрывном единстве:

$$H = H_n + H_\tau + H_L, \quad (7-48)$$

где H_n , H_τ и H_L берутся согласно (7-45а), (7-46а) и (7-47а).

Это значит, что истинная суть H объективной реальности складывается в каждый момент времени из сформировавшегося к этому моменту содержимого памяти (отраженной сути), из той логической информации H_τ , которая пребывает в процессе передачи от органов чувств к памяти, и из той логической информации H_L , усвоению которой препятствуют привычки и предрассудки. Понятно, что две последние составляющие преходящи и в конце концов (при $t = \infty$) исчезают, оставляя в памяти адекватное логическое отражение объективной реальности $H_n = H$.

Наконец, последний закон диалектики, подлежащий учету, — закон перехода количественных изменений в коренные качественные — требует понимания того, что характеристические константы понятий n , τ и L являются константами лишь в ограниченном диапазоне эволюционных изменений понятия. В общем же случае, т. е. при больших изменениях, эти константы могут измениться, что приведет к радикальному (революционному) изменению сути понятия, нарушив плавный (эволюционный) ход изменений. С формальной точки зрения это означает, что в общем случае n , τ и L являются функциями H , J , dJ/dt и d^2J/dt^2 .

Отметим, что хотя n , τ и L мы называли соответственно емкостью, сопротивлением и ригидностью понятия, но поскольку последние формируются в сознании конкретного человека, то эти параметры, по существу, характеризуют память, пропускную способность и ригидность психики этого человека и могут быть непосредственно измерены при эксперименте с участием человека.

С учетом всего вышеизложенного (7-48) символизируют грамотное суждение о том или ином элементе системы, или, другими словами, дает исчерпывающее определение этого элемента.

Настало время обратить теперь внимание на то, что операторы d/dt и d^2/dt^2 , символизирующие изменения, ни по форме, ни по содержанию не отличаются от математических символов дифференцирования, так что принятое нами символическое изображение суждения или определения понятия является, по существу, дифференциальным уравнением. Это означает, в свою очередь, что процедура преобразования суждений, записанных в вышеприведенной символической форме, ничем не отличается от принятой в математике процедуры преобразования и решения дифференциальных уравнений. Иными словами, сконструированная символическая методология, т. е. формальная диалектическая логика, оказалась, по существу, дифференциальным исчислением, которое и есть символическая форма диалектической логики. Действительно, если рассматривать несколько символических суждений (7-48) по какому-либо поводу как систему уравнений, то решение ее возможно лишь в том случае, когда число суждений будет не меньше числа объектов, о которых должно быть сделано умозаключение, что соответствует логическому закону достаточного основания. Кроме того, для решения системы уравнений необходимо, чтобы все уравнения были совместны друг с другом, что имеет место, когда определитель Δ системы не равен нулю:

$$\Delta = \begin{vmatrix} 1/n_{11} & 1/n_{12} & \dots & 1/n_{1m} \\ 1/n_{21} & 1/n_{22} & \dots & 1/n_{2m} \\ \dots & \dots & \dots & \dots \\ 1/n_{m1} & 1/n_{m2} & \dots & 1/n_{mm} \end{vmatrix} \neq 0.$$

Это соответствует логическому закону противоречия, требующему непротиворечивости (совместности) суждений.

Наконец, логический закон тождества выражается в естественном требовании изображать синонимы одним и тем же символом.

Впрочем, из правил математики вытекает и еще один логический закон, согласно которому тавтологии (линейные комбинации суждений) не изменяют исходной системы и обнаруживаются по признаку $\Delta = 0$ и $\Delta_k = 0$, где Δ_k отличается от Δ заменой в одном столбце $1/n_{ik}$ на H_i .

Необходимо теперь рассмотреть, каким образом наши символические суждения (7-45)–(7-48) соотносятся с принятыми в физике дифференциальными уравнениями процессов. Обратимся, например, к дифференциальному уравнению электрической цепи, состоящей из последовательно соединенных конденсатора, резистора и катушки индуктивности:

$$u = q/c + rdq/dt + L_3 d^2q/dt^2.$$

Это уравнение, по существу, символически изображает суждение о том, что напряжение u цепи складывается: 1) из напряжения, создаваемого зарядом q конденсатора, имеющего емкость c ; 2) напряжения, создаваемого электрическим током dq/dt на резисторе r ; 3) напряжения, создаваемого изменением тока d^2q/dt^2 на катушке индуктивности L_3 . Обозначив $J = q/\Delta q$ и $H = u/\Delta u$, где Δq и Δu — «кванты» соответственно заряда и напряжения, которые дальше делить нецелесообразно, получаем из уравнения электрической цепи символическое суждение (7-48), в котором $n = c\Delta u/\Delta q$, $\tau = r\Delta q/\Delta u$ и $L = L_3\Delta q/\Delta u$ и которое не только по существу, но и по форме символизирует суждение о сущности понятия «напряжение электрической цепи», определяя его через статическую, кинетическую и динамическую составляющие, как того требует диалектика.

Таким образом, посредством квантования переменных в любом дифференциальном уравнении нетрудно перейти от него к символическим суждениям диалектической логики и наоборот.

Всякое понятие, как отмечалось, всегда есть относительная истина, характеризующаяся определенной достоверностью. Между тем, если достоверность g понятия может меняться в пределах от 0 до 100%, то сущность H понятия согласно, например, (7-45) может меняться в пределах от нуля до бесконечности, поэтому, чтобы установить однозначное соответствие между ними, придется прибегнуть к преобра-

зованию вида

$$H = -\log(1 - g), \quad (7-49)$$

которое по форме аналогично информации Шеннона, что и позволяет измерять H в битах, если принять в (7-49) двоичный логарифм. Подчеркнем, что в (7-49) достоверность g понятия характеризует вероятность p достижения той или иной цели при использовании данного понятия. Следовательно, чем достовернее понятие, тем успешнее используется оно в практической деятельности, и, наоборот, чем успешнее практическая деятельность, тем выше достоверность используемых понятий. Однако, поскольку одно и то же понятие при его использовании в разного рода деятельности дает неодинаковую вероятность достижения цели, то достоверность g и вероятность P достижения цели не тождественны, а лишь пропорциональны друг другу:

$$g = qP, \quad (7-50)$$

где q — степень принадлежности понятия данной деятельности, т. е. доля этой деятельности в формировании понятия. Обычно $0 \leq q \leq 1$.

Если понятие нигде, кроме данного рода деятельности, не используется, то оно полностью принадлежит этой деятельности, так что $q = 1$ и $g = P$. Но если, напротив, понятие не имеет к данному роду деятельности никакого отношения, то независимо от успеха или неуспеха в достижении цели невозможно судить о достоверности этого понятия, поскольку $q = 0$.

Вернемся теперь к логическому закону исключенного третьего и покажем, что наше первоначальное предположение о непримиримости его в диалектической логике верно лишь отчасти. Действительно, в традиционной формулировке «Из двух противоположных суждений всегда одно истинно, другое ложно, а третьего быть не может» этот закон неприменим к относительно истинным противоположным суждениям, поскольку они могут быть оба истинными в той или иной мере. Например, противоположные суждения «Мул — лошадь» и «Мул — не лошадь» оба резонны, поскольку мул — гибрид лошади и осла, соединяющий в себе признаки обоих родителей. Здесь оба суждения сливаются, по существу, в одно: «Мул — лошадь и не лошадь», в котором мул выражает неразрывное, диалектическое единство противоположностей. При этом, если достоверность первого суждения g_1 , а достоверность вто-

рого g_2 , то всегда имеет место диалектический закон единства противоположностей:

$$g = g_1 + g_2 = 1, \quad (7-51)$$

который гласит, что совокупность противоположных суждений всегда абсолютно истинна.

Это, разумеется, справедливо и в том случае, когда достоверность одного из противоположных суждений равна единице, а другого — нулю, т. е. когда одно суждение абсолютно истинно, а другое абсолютно ложно, что составляет традиционную область действия закона исключенного третьего. Таким образом, диалектический закон (7-51) единства

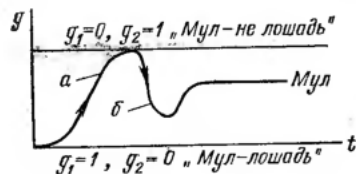


Рис. 7-15. К закону отрицания отрицания

ва противоположностей представляет собой всегда справедливое обобщение логического закона исключенного третьего, в котором (обобщении) последний выступает как предельный частный случай. Кроме того, из (7-49) и (7-51) следует, что суть совокупности противоположных суждений всегда равна бесконечности: $H = -\log 0 = \infty$. Но это же следует и из закона противоречия для несовместных суждений, для которых $\Delta = 0$, что лишний раз подтверждает синонимичность, фактическое совпадение законов противоречия и исключенного третьего, а также невозможность умозаключения на основе полностью противоположных суждений. Это, однако, не только не запрещает, но даже требует для полноценного умозаключения наличия диалектического противоречия в суждениях, достоверность которых отлична от нуля и единицы и удовлетворяет (7-51), т. е. диалектическому закону противоречия, причем сами суждения выступают как границы, в пределах которых заключена относительная истина (рис. 7-15).

Если на графике откладывать по оси ординат достоверность суждений, например, о муле («Мул — лошадь» — g_1 , «Мул — не лошадь» — g_2), то границами области формирования понятия о муле будут горизонтальные прямые, для одной из которых $g_1 = 1$ и $g_2 = 0$, а для другой $g_1 = 0$ и $g_2 = 1$, а самому понятию «мул» будет соответствовать некое промежуточное состояние. Поскольку понятие формируется во времени t (ось абсцисс), закон единства противоположностей всегда имеет в виду последовательное во

времени формирование в сознании противоречивых суждений, что нисколько не мешает соответствующим противоположностям в объективной реальности существовать в один и тот же момент времени («Мул — одновременно лошадь и не лошадь»). Это значит, что если вначале в нашем сознании складывается тезис «Мул — лошадь», то после сопоставления его с действительностью должен возникнуть антитезис «Мул — не лошадь», что графически (рис. 7-15) представляется как путь в направлении стрелки от одной границы к другой. Поскольку процесс сопоставления признаков мула с признаками лошади сам протекает последовательно во времени, то этот путь представляется какой-то плавной кривой (а).

Следующий этап формирования понятия должен быть связан с исследованием другой границы, т. е. с сопоставлением мула с признаками не лошади, которое является сферой действия закона отрицания отрицания, ибо после всех сопоставлений должно иметь место отрицание и другой границы, т. е. синтез понятия «не лошадь». Но на графике «не лошадь» означает движение от границы «Мул — не лошадь» в сторону границы «Мул — лошадь», т. е. в обратном направлении (б). Вместе с тем в общем случае возврат на саму границу бессмыслен, поскольку она уже отрицалась в антитезисе, так что остановка должна произойти в той или иной промежуточной точке. При этом не исключено повторное и даже многократное применение закона отрицания отрицания, что графически будет изображаться колебательным процессом, постепенно сходящим к установившемуся значению понятия «мул», с конечной достоверностью как тезиса, так и антитезиса в рамках (7-51), причем этот процесс должен определяться индивидуальными особенностями психики, т. е. памятью n , сопротивлением (сообразительностью) τ и ригидностью L индивида.

Действительно, решение дифференциального уравнения (7-48) для содержимого памяти H_n часто дает затухающую синусоиду с периодом $T = \sqrt{nL}$ и затуханием $\delta = \frac{\tau}{2} \sqrt{n/L}$:

$$H_n = H \left[1 - \left(\cos \frac{t\sqrt{1-\delta^2}}{T} + \frac{\delta}{\sqrt{1-\delta^2}} \sin \frac{t\sqrt{1-\delta^2}}{T} \right) \times \exp(-\delta t/T) \right],$$

представляющую собой согласно (7-49) логарифм кривой, изображенной на рис. 7-15. Это выражение описывает

процесс формирования понятия во времени, т. е. процесс накопления в памяти человека соответствующей логической информации H_n , отражающей суть H соответствующей объективной реальности. Согласно последнему выражению логическое отражение сути явления адекватно сути реальности ($H_n = H$) только в конце процесса при $t \rightarrow \infty$ (кривая 1 на рис. 7-16). В остальное время мы имеем дело лишь с относительной истиной ($H_n \neq H$).

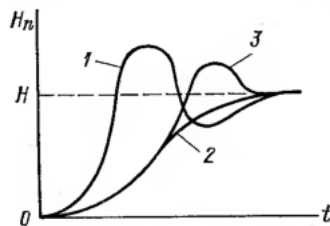


Рис. 7-16. К формированию понятия устойчивой психикой

Отметим, что процесс формирования понятия носит колебательный характер только при $\delta < 1$, т. е. когда имеет место следующее сочетание параметров психики: $\pi\tau^2 < 4L$.

У людей, имеющих такое сочетание параметров психики, понятия формируются относительно быстро, но сопровождаются сомнениями и перепроверками, т. е. медленно превращаются в убеждения. Если же $\delta > 1$ и $\pi\tau^2 > 4L$, то

понятия формируются относительно быстро, но сопровождаются сомнениями и перепроверками, т. е. медленно превращаются в убеждения. Если же $\delta > 1$ и $\pi\tau^2 > 4L$, то

$$H_n = H \left[1 - \left(\operatorname{ch} \frac{t\sqrt{\delta^2 - 1}}{T} + \frac{\delta}{\sqrt{\delta^2 - 1}} \operatorname{sh} \frac{t\sqrt{\delta^2 - 1}}{T} \right) \times \exp(-\delta t/T) \right],$$

т. е. процесс превращения незнания ($H_n = 0$) в знание ($H_n = H$) протекает плавно, хотя и относительно медленно, и сразу в форме твердого убеждения (кривая 2 на рис. 7-16). Оптимальным с точки зрения формирования понятия, видимо, является промежуточное между приведенными сочетание параметров психики $\pi\tau^2 = 4L$ и $\delta = 1$, когда

$$H_n = H [1 - (1 + t/T) \exp(-t/T)]$$

(кривая 3 на рис. 7-16), т. е. когда понятие формируется быстрее всего.

Вместе с тем кривая 3 и особенно кривая 1 вначале быстро достигают штриховой горизонтали, обозначающей суть H объективной реальности, затем пересекают ее, не задерживаясь, что, видимо, и дает основание для широкого распространения убеждения, что первое впечатление самое верное.

Приведенные решения уравнения (7-48) относятся только к случаю формирования понятия в условиях неизменной

сущности ($H = \text{const}$) отражаемого объекта. Если же формирование понятия происходит медленно, а объект эволюционирует относительно быстро, то его суть является функцией времени $H(t)$ и решение уравнения может иметь другой вид. Кроме того, согласно диалектическому закону перехода количественных изменений в качественные в приведенных решениях параметры T и δ могут сами оказаться функциями H и t , что требует соответствующего дополнительного исследования параметров психики n , τ и L . Наконец, излагаемая символическая логика в сильной степени зависит от того, склонна данная индивидуальная психика к рутинному или к творческому мышлению.

Рутинное мышление способно лишь к пассивному отражению заданной сути вещей, что следует из приведенных решений для $\delta > 0$. Что касается творческого мышления, то для него заданная суть является лишь толчком для непрерывного и безграничного творческого процесса, которому соответствует $\delta < 0$. При этом, если $\delta > -1$, то процесс носит характер расходящихся колебаний бесконечно возрастающей амплитуды вокруг заданной сущности с бесконечным исследованием всех альтернатив возможного развития объективной реальности (кривая 1 на рис. 7-17). Если же $\delta < -1$, то процесс творчества носит бескомпромиссно целеустремленный, но относительно неторопливый характер (кривая 2 на рис. 7-17), направленный на безграничное усовершенствование объективной реальности. Наконец, если $\delta = -1$, то имеет место самое быстрое творчество (кривая 3 на рис. 7-17).

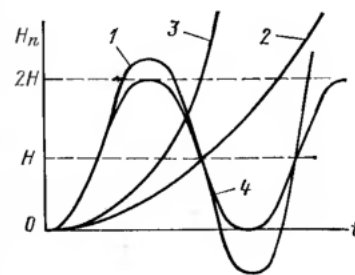


Рис. 7-17. К формированию понятия при творческой деятельности

Можно выделить и еще один граничный между рутинным и творческим процесс (кривая 4 на рис. 7-17), когда $\delta = 0$. Этот процесс сводится к бесконечному исследованию возможных альтернатив в ограниченной области от нуля до $2H$ и характерен для неустойчивой психики, по существу, не способной ни к рутинной, ни к творческой деятельности и лишенной каких-либо стабильных убеждений.

Таким образом, в принципе рутинный и творческий интеллекты представляют собой две крайние противоположности, разделенные переходной областью существования сравнительно малороботоспособного неустойчивого интеллекта. Однако с учетом нелинейности параметров реальной психики не исключен ее переход от рутинного к творческому мышлению и обратно через неработоспособное состояние в случае значительных изменений H и стрессовых воздействий.

Это значит, что применение диалектической логики к исследованию систем всегда связано с субъективными оценками, проистекающими из индивидуальных особенностей психики исследователя. С учетом этого обстоятельства изложенный формализм применим к исследованию систем любого рода, в том числе и человекомашинных систем.

7.4. АНАЛИЗ ИЕРАРХИЧЕСКИХ СТРУКТУР

Обратимся теперь к широко распространенным древовидным иерархическим системам управления и сбора информации. Рассмотрим, какие возможности и способы исследования предоставляет информационный подход в этом случае, для чего введем ряд дополнительных новых понятий и определений.

Сложность системы из J одинаковых независимых элементов, т. е. таких, состояния каждого из которых не зависят от состояний других элементов, характеризуется числом возможных состояний системы как целого, которое в данном случае, очевидно, составляет K^J , где K — число равновероятных состояний одного элемента. Однако данная оценка сложности становится неудобной, как только речь пойдет о системах из неодинаковых элементов или из элементов с неравновероятными состояниями. Поэтому более удобной оказывается оценка сложности, которая получается из вышеприведенной посредством логарифмирования:

$$C = \log K^J = J \log K = -J \log P_k, \quad (7-52)$$

где $P_k = 1/K$ — вероятность k -го состояния элемента.

Обобщая (7-52) на случай применения элементов с неравновероятными состояниями, получаем:

$$C = -J \sum_{k=1}^K P_k \log P_k = JH, \quad (7-53)$$

где H — энтропия элемента, по Шеннону [54], отражающая его сущность в системе.

Наконец, в общем случае для применения различных элементов

$$C = \sum_{i=1}^J H_i = \sum_{j=1}^m P_j \log P_j, \quad (7-54)$$

где m — число состояний системы как целого; P_j — вероятность j -го состояния системы.

В дальнейшем полученную оценку C будем именовать содержанием системы, H_i — сущностью элемента, а J — информацией.

Из (7-52) следует, что, поскольку в $H = \log K$ минимальное отличное от нуля значение H соответствует числу K состояний элемента, равному двум, то это значение H целесообразно принять в качестве единицы измерения («кванта») сущности. Тогда основание логарифма придется принять равным двум, чтобы обеспечить $\log 2 = 1$. Эта единица измерения сущности называется бит и используется в дальнейшем. Итак, $1 \text{ бит} = \log_2 2$.

Нетрудно заключить, что не только сущность H , но и информация J измеряется в битах, поскольку всегда справедливо тождество $J = \log_2 2^J = J \log_2 2 = J \text{ бит}$. Тогда содержание системы согласно (7-53) должно измеряться в бит². В дальнейшем мы всегда будем пользоваться этими единицами, опуская в формулах число два — основание логарифмов.

Таким образом, согласно (7-52) содержание системы из трех независимых элементов с двумя равновероятными состояниями у каждого составляет $C_1 = 3 \log 2 = 3 \text{ бит}^2$, содержание элемента с восемью состояниями также составляет $C_2 = C_1 = 1 \log 8 = 3 \text{ бит}^2$, а содержание системы из двух независимых элементов с тремя состояниями несколько больше и составляет $C_3 = 2 \log 3 = 3,2 \text{ бит}^2$.

Практический интерес представляет определение системы с максимальной сущностью элементов, т. е. с максимальным содержанием, приходящимся на элемент системы. Ограничимся рассмотрением только таких систем, в которых число состояний системы равно числу элементов, из которых она состоит, т. е. равно информации J .

В классе таких систем имеем согласно (7-53) $H_k = \log J/J$. Максимум сущности элемента соответствует $dH_k/dJ = (1/J^2) \log(e/J) = 0$, откуда следует $J = e$ и $H_{k \text{ макс}} = 0,535 \text{ бит}$.

Таким образом, максимальной сущностью элементов в классе рассматриваемых систем обладала бы система с числом элементов, равным числу Эйлера, ближайшим целым к которому является число 3, чем, вероятно, объясняется тяготение человека ко всякого рода триадам и троицам. При этом, хотя число 2 дальше от числа Эйлера, чем число 3, тем не менее для системы из двух элементов $H_k = 0,5$ бит, т. е. мало отличается от сущности элементов в трехэлементной системе, где $H_k = 0,53$ бит. Поскольку, однако, двухэлементная система проще трехэлементной, в природе и в человеческой деятельности двухэлементные системы распространены едва ли не больше, чем трехэлементные.

До сих пор мы рассматривали только так называемое системное содержание, которому в (7-54) отвечает P_j , представляющее собой совместную вероятность соответствующих состояний элементов системы, т. е. произведение условных вероятностей тех или иных состояний различных элементов. Между тем, помимо системного содержания C_c , можно выделить [17] еще собственное содержание C_0 системы (которому в (7-54) соответствует P_j , представляющее собой априорную совместную вероятность отдельных состояний различных элементов системы, т. е. произведение вероятностей отдельных априорных состояний элементов вне связи их между собой) и взаимное содержание $C_{вз}$ системы элементов, представляющее собой разность системного и собственного содержаний

$$C_{вз} = C_c - C_0 \quad (7-55)$$

и соответствующее обобщению энтропии, введенной Колмогоровым [27]. В качестве примера продолжим рассмотрение трехэлементной системы с тремя равновероятными состояниями ($P_j = 1/3$), имеющей максимально достижимую системную сущность дискретных элементов. Для этой системы согласно (7-54) системное содержание составляет $C_c = \log 3 = 1,6$ бит². При этом каждый из элементов системы может, вообще говоря, иметь сколько угодно состояний, лишь бы эти состояния были так заблокированы, что в совокупности обеспечивали всего три состояния системы как целого. Если же система состоит из двухпозиционных элементов, в каждом из которых оба состояния равновероятны, то их априорная совместная вероятность составит $P_j = (0,5)^3$, чему согласно (7-54) соответствует собственное содержание системы $C_0 = \log 8 = 3$ бит².

Таким образом, взаимное содержание согласно (7-55) составит $C_{вз} = -1,4$ бит².

Интересно отметить, что собственное содержание C_0 такой оптимальной по сущности двоичных элементов системы, характерной для нервной системы человека, соответствует равновероятному выбору из восьми, что подтверждает гипотезу Миллера [32], согласно которой оперативная память человека способна оперировать в среднем только с семью объектами. Следует лишь оговорить, что в приводимых Миллером случаях тяготения человека к семи дням недели, семи цветам спектра, семи чудесам света и т. д. не учитывается всегда имеющее место восьмое состояние, соответствующее отсутствию всех перечисленных качеств, т. е. когда день не рассматривается как часть недели (например, праздник) и т. п.

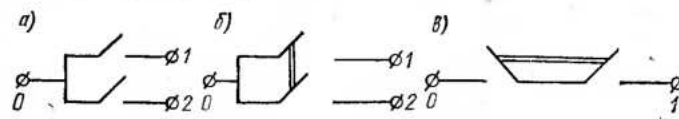


Рис. 7-18. К целостности систем

Явление, отмеченное гипотезой Миллера, может быть одной из причин структуризации (представления в виде структуры) чисто человеческих систем управления, поскольку указывает на ограниченные возможности оперативной памяти одного человека в процессе управления.

Об анализе структуры систем. Вышерассмотренный информационный аппарат позволяет проводить анализ структуры материальных систем любой физической природы. При этом *первый вопрос*, который необходимо решить, — это вопрос о том, составляет ли изучаемая совокупность элементов целое, т. е. является ли она действительно системой или только создает обманчивую видимость системы. Так, например, схема на рис. 7-18, а не составляет единого целого, хотя и создает его видимость, а схемы на рис. 7-18, б и 7-18, в составляют целое. Решить вопрос о целостности можно, исследуя содержание системы. При этом, если система, помимо собственного содержания, обладает еще и взаимным содержанием, то она имеет свойство целостности. Если же система не обладает взаимным содержанием, то она не имеет свойства целостности. Взаимное содержание проще всего определить согласно (7-55) как разность системного и собственного содержания системы. Обратим внима-

ние читателей на то обстоятельство, что из формулы (7-54), если вести суммирование по элементарным носителям, каждый из которых содержит 1 бит информации, следует, что системное содержание численно равно системной сущности, а собственное содержание численно равно сумме сущностей отдельно взятых элементов. Таким образом, схема на рис. 7-18, а, имеющая четыре равновероятных состояния, соответствующих соединению вывода 0 с выводом 1, 0 с 2, 0 с 1 и 2 одновременно и, наконец, отсутствию всякого соединения между ними, обладает сущностью $H_c = -\log 0,25$, т. е. системным содержанием $C_c = 2 \text{ бит}^2$. В то же время схема на рис. 7-18, б, имеющая всего два равновероятных состояния, соответствующих замыканию между собой всех выводов либо их полному размыканию, обладает сущностью $H_c = -\log 0,5$, т. е. системным содержанием $C_c = 1 \text{ бит}^2$. В свою очередь, схема на рис. 7-18, в имеет всего одно только разомкнутое состояние, т. е. ее системное содержание равно нулю. Все эти схемы состоят из двух элементов (ключей), собственное содержание которых при равновероятных состояниях составляет согласно (7-52) $C_0 = 2 \text{ бит}^2$. Таким образом, для схемы а взаимное содержание ключей согласно (7-55) равно нулю и она не обладает свойством целостности; для схемы б взаимное содержание ключей составляет $C_{вз} = -1 \text{ бит}^2$ и эта схема обладает свойством целостности, как и схема в, для которой $C_{вз} = -C_0 = -2 \text{ бит}^2$.

Оценка содержания частей и целого позволяет сделать вывод, что сутью всякого устойчивого объединения является уменьшение содержания целого по сравнению с собственным содержанием частей за счет установления их взаимозависимости, т. е. за счет появления взаимного содержания частей, являющегося смыслом целого. Например, если мы имеем набор из нескольких букв, который не составляет для читателя понятного целого, то содержание этого сочетания равно сумме содержаний отдельных букв, поскольку мы здесь имеем дело с последовательным соединением букв, при котором их сущности H суммируются. Таким образом, в этом случае $C_0 = J_1 H$, где J_1 — число букв в наборе.

Если же упомянутый набор букв составляет понятное целое, то буквы в наборе становятся взаимосвязанными (избыточными), что позволяет нам прибегать к сокращениям понятных слов. Если мы используем безыбыточное сокращение слова, т. е. такое слово, которое уже нельзя сокра-

тить без ущерба для понимания, содержащее только $J_1 < J_2$ букв из всего слова, то $C_c = J_2 H < C_0$.

Такая же картина наблюдается при объединении слов во фразы. При этом системное содержание всей фразы, хотя и меньше суммарного содержания слов, но не может быть меньше собственного содержания одного слова. Этот предельный случай имеет место во фразе «масло масляное», где содержание всей фразы равно содержанию одного слова. Если же слова во фразе вступают в противоречие, то системное содержание неустойчивой фразы должно быть больше собственного содержания слов. Например, в словосочетании «круглый квадрат» каждое из слов имеет вполне определенное содержание, а вместе они имеют бесконечное содержание, поскольку могут означать все что угодно.

Вторым вопросом, который возникает при анализе систем, является вопрос о степени целостности, или просто о целостности α как числовой характеристике системы. Для схемы на рис. 7-18 можно сделать заключение о целостности по значению $C_{вз}$, полагая, что чем больше взаимное содержание, тем больше целостность системы. Однако такая возможность проистекает из того обстоятельства, что все три схемы (рис. 7-18) имеют одинаковое собственное содержание элементов C_0 . Для систем с различным значением C_0 мало что дает сопоставление взаимных содержаний. В таком случае гораздо удобнее использовать относительную оценку $\alpha = -C_{вз}/C_0$, которая для вышеассмотренных схем составляет соответственно $\alpha_a = 0$, $\alpha_b = 0,5$, $\alpha_v = 1$. При этом нулевая целостность означает простую совокупность несвязанных частей, а целостность, равная единице, — абсолютную целостность. Иногда целесообразна сопряженная оценка $\beta = C_c/C_0 = 1 - \alpha$, которую можно назвать коэффициентом использования частей в целом. Применительно к рассматриваемым схемам эта оценка дает соответственно $\beta_a = 1$, $\beta_b = 0,5$, $\beta_v = 0$. Иными словами, не обладающая целостностью система элементов подразумевает наиболее универсальное, полное использование свойств своих частей, а абсолютно целостная система, напротив, вообще лишает элементы их первоначальных свойств, используя лишь те свойства, которые присущи системе как целому и которые не содержатся в отдельно взятых ее элементах. Например, сравнивая между собой схемы а и б на рис. 7-19, осуществляющие равновероятный выбор из восьми положений, чему соответствует $C_c = 3 \text{ бит}^2$, можно заключить, что для схемы а $C_0 = 3 \text{ бит}^2$, $C_{вз} = 0$,

$\beta = 1$, а для схемы b $C_0 = 7 \text{ бит}^2$, $C_{\text{вз}} = -4 \text{ бит}^2$, $\beta = 3/7$, т. е. что использование ключей в схеме b в среднем хуже, чем в схеме a , хотя эти схемы имеют одинаковое системное содержание, определяемое одинаковым их назначением.

Однако среднее значение коэффициентов использования элементов не позволяет судить о реальных условиях работы того или иного из них в тех случаях, когда эти условия различны для разных элементов, поэтому *третий вопрос*, который необходимо решить при анализе системы, — это вопрос о коэффициенте использования каждого из ее элементов. Решение этого вопроса особенно актуально для

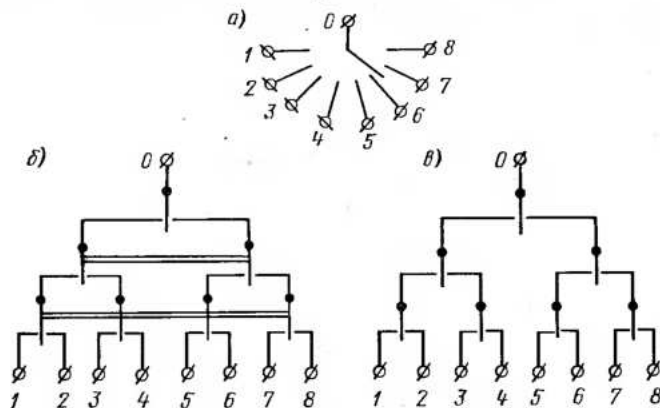


Рис. 7-19. К выбору из восьми состояний

иерархических структур, элементы которых почти всегда имеют отличное друг от друга содержание в зависимости от уровня иерархии, которому они принадлежат. Так, иерархическая схема на рис. 7-19, b содержит три уровня иерархии, каждый из которых имеет два равновероятных состояния, чему соответствует системное содержание $C_s = 1 \text{ бит}^2$. Однако на верхнем уровне все это содержание принадлежит единственному ключу, для которого, таким образом, $C_s = C_0 = 1 \text{ бит}^2$, $\beta = 1$. На среднем уровне имеется два равноправных ключа, между которыми поровну делится содержание уровня, так что для каждого из них $C_s = 0,5 \text{ бит}^2$, $C_0 = 1 \text{ бит}^2$, $C_{\text{вз}} = -0,5 \text{ бит}^2$, $\beta = 0,5$. Наконец, на нижнем уровне его содержание делится между четырьмя равноправными ключами, так что для каждого из них $C_s = 0,25 \text{ бит}^2$, $C_0 = 1 \text{ бит}^2$, $C_{\text{вз}} = -0,75 \text{ бит}^2$, $\beta = 0,25$.

Из сказанного следует, что только ключ верхнего уровня используется полностью, ключи среднего уровня — наполовину, а ключи нижнего уровня — только на четверть их собственного содержания. Оставшаяся часть собственного содержания приходится на их взаимное содержание, т. е. на содержание взаимодействия ключей в пределах своего уровня.

Все рассмотренные схемы подразумевали использование всех их возможностей для достижения тех или иных целей, так что мы не делали различия между их системным содержанием и смыслом, поскольку в этом случае они количественно совпадают. Между тем иногда прикладное содержание (смысл) системы бывает уже, чем ее полное содержание. Это имеет место, когда для достижения определенных целей используется только часть потенциальных возможностей системы (орудия). Поэтому *четвертый вопрос*, возникающий при анализе структуры системы, — это вопрос о коэффициенте полезного действия (КПД) как отдельных элементов, так и системы в целом.

Под КПД γ понимается при этом отношение смысла C элемента или системы к системному содержанию, т. е. $\gamma = C/C_s$. Например, если схема на рис. 7-19, a , имеющая содержание $C_s = C_0 = 7 \text{ бит}^2$ и $\beta = 1$, используется только для соединения вывода 0 с одним из выводов 1—8, т. е. только для равновероятного выбора из восьми, то ее смысл составляет $C = 3 \text{ бит}^2$, а $\gamma = 3/7$, что меньше КПД схем a и b , у которых он равен единице. Что касается КПД ключей отдельных уровней, то они равны единице для всех ключей схем a и b . Для схемы b КПД ключа верхнего уровня составляет $\gamma = 1$, ключей среднего уровня $\gamma = 0,5$ и ключей нижнего уровня $\gamma = 0,25$, поскольку этот уровень вместо необходимого одного выбора из двух делает четыре таких же выбора.

Таким образом, сравнение аналогичных по назначению схем b и a (рис. 7-19) показывает, что схема a обладает большими возможностями, чем схема b .

Сравним теперь по КПД схемы на рис. 7-20, символизирующие различные варианты иерархической структуризации задачи равновероятного выбора из восьми состояний, смысл которой составляет $C = \log 8 = 3 \text{ бит}^2$. При этом схема a символизирует неструктуризованную задачу прямого выбора, для которой $C = C_s = C_0 = \log 8 = 3 \text{ бит}^2$, $\beta = 1$, $\gamma = 1$; схема b символизирует однородную дихотомическую трехуровневую иерархическую структуру, для

которой $C_c = C_0 = 7 \log 2 = 7 \text{ бит}^2$, $\beta = 1$, $\gamma = 3/7$; схемы *в* и *г* символизируют двухуровневые, неоднородные по уровням иерархии, причем для *в* $C_c = C_0 = \log 2 + 2 \log 4 = 5 \text{ бит}^2$, $\beta = 1$, $\gamma = 0,6$, а для *г* $C_c = C_0 = \log 4 + 4 \log 2 = 6 \text{ бит}^2$, $\beta = 1$, $\gamma = 0,5$; схема *д* символизирует почти однородную двухуровневую иерархию, для которой $C_c = (-0,375 \log 0,375 - 0,25 \log 0,25 - 0,375 \log 0,375) + 2 \log 3 + \log 2 = 5,75 \text{ бит}^2$, $C_0 = 3 \log 3 + \log 2 = 5,8 \text{ бит}^2$, $\beta = 0,99$, $\gamma = 0,52$, причем

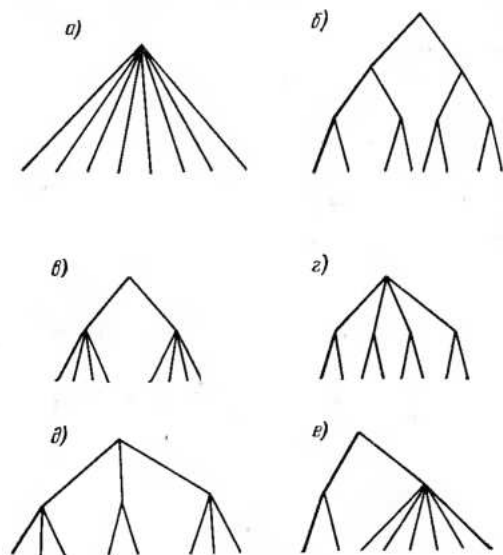


Рис. 7-20. Иерархические структуры с различным числом уровней

выражение в скобках означает сущность элемента верхнего уровня, имеющего три неравновероятных состояния; наконец, схема *е* символизирует неоднородную как по уровням, так и в пределах уровня несимметричную иерархию, для которой $C_c = (-0,25 \log 0,25 - 0,75 \log 0,75) + \log 2 + \log 6 = 4,4 \text{ бит}^2$, $C_0 = 2 \log 2 + \log 6 = 4,6 \text{ бит}^2$, $\beta = 0,96$, $\gamma = 0,68$.

Таким образом, хуже других по коэффициенту использования своих элементов схемы *д* и *е*, неоднородные в пределах одного уровня, а хуже других по КПД схемы *б* и *г*, содержащие наибольшее число элементов. Абсолютно луч-

шей по всем этим показателям является неструктуризованная одноэлементная схема *а*, так что чем сильнее структуризована задача и чем неоднороднее структура в пределах каждого уровня, тем хуже ее показатели γ и β .

Но тогда возникает вопрос: в чем же смысл структуризации? Чтобы ответить на него, необходимо обратиться к смысловой нагрузке элементов структуры и к определению затрат умственного труда на решение неструктуризованной и структуризованной задач.

Причины структуризации. Что касается смысловой нагрузки схемы, то вполне очевидно, что она делится между отдельными элементами иерархии; позволяя тем самым выполнять задачу по частям либо последовательно во времени одним элементом (человеком, автоматом), либо параллельно несколькими элементами, которые в этом случае осуществляют разделение труда, сотрудничество, кооперацию. Если элементы взаимно дополняют друг друга в процессе решения задачи (достижения цели), то такая система обладает устойчивостью, о чем свидетельствует ее положительная целостность ($\alpha > 0$). Если же элементы находятся в противоречивых отношениях, т. е. мешают друг другу в решении задачи, то такая система неустойчива, о чем свидетельствует ее отрицательная целостность ($\alpha < 0$). Например, внутренняя энергия (масса) устойчивых ядер атомов, являющаяся системным содержанием физического взаимодействия, меньше внутренней энергии (массы) составляющих ядро протонов и электронов, являющейся их собственным физическим содержанием, так что взаимное их содержание в ядре согласно (7-55) отрицательно, а целостность ядра положительна: $\alpha = -C_{вз}/C_0 > 0$. Напротив, внутренняя энергия неустойчивых (радиоактивных) ядер больше внутренней энергии составляющих ядро частиц, так что $C_{вз} > 0$, а целостность структуры ядра отрицательна: $\alpha = -C_{вз}/C_0 < 0$.

Рассмотрим теперь смысловую нагрузку элементов устойчивых структур на рис. 7-20. В случае неструктуризованной задачи (рис. 7-20, *а*) вся ее смысловая нагрузка $C = 3 \text{ бит}^2$ приходится на один элемент, так что если под рукой нет элемента (человека, автомата), способного в одиночку справиться с задачей, ее приходится структуризовать, т. е. разделять на части, решение которых под силу наличным элементам. Если мы располагаем только двоичными элементами, каждый из которых способен вынести за один раз смысловую нагрузку, равную его содержанию, т. е.

$C_k = C_{0k} = \log 2 = 1$ бит², то получим структуру б, число уровней в которой определяется отношением $C/C_k = 3$. Однако в этих условиях лишь элемент верхнего уровня работает в каждом акте выбора; остальные же элементы работают либо через раз (средний уровень), либо через три раза (нижний уровень), поэтому для полной характеристики условий их работы необходимо учесть время, например, в форме

$$N = dC/dt. \quad (7-56)$$

Здесь N — смысловая мощность, т. е. смысл, перерабатываемый элементом или системой в единицу времени.

Согласно (7-56) смысловая мощность элемента верхнего уровня схемы б в два раза больше смысловой мощности элемента среднего уровня и в четыре раза больше смысловой мощности элемента нижнего уровня, так что, если мы хотим поместить элементы всех уровней в одинаковые условия, нам придется использовать на среднем уровне элементы, обладающие в два раза большим содержанием, чем на верхнем уровне, т. е. элементы с четырьмя равновероятными состояниями, у которых $C_k = \log 4 = 2$ бит². Но тогда приходим к структуре в, где для всех элементов N одинаково. Это тем более целесообразно, что и КПД схемы в выше, чем КПД схемы б. Итак, если на решение задачи отведено время T , то согласно (7-56) схема (элемент) а должна развить $N = 3/T$ бит²/с, а в схеме в элементы развивают в 3 раза меньшую смысловую мощность ($N = 1/T$ бит²/с), так что польза структуризации здесь очевидна, ибо задача решается усилиями маломощных (менее способных, хотя и нескольких) элементов.

Понятно, что в этих условиях схемы г—е должны быть хуже схемы в. Действительно, для схемы г на верхнем уровне $N = 2/T$, а на нижнем $N = 1/(4T)$, т. е. мощность в 8 раз меньше; для схемы д на верхнем уровне $N = (-0,375 \log 0,375 - 0,25 \log 0,25 - 0,375 \log 0,375)/T = 1,55/T$, а на нижнем уровне $N_2 = 0,25/T$ и $N_3 = 0,375 \log 3/T = 0,6/T$, т. е. двоичный элемент несет в 6 раз меньшую нагрузку, чем элемент верхнего уровня; наконец, для схемы е на верхнем уровне $N = (-0,25 \log 0,25 - 0,75 \log 0,75)/T = 0,8/T$, а на нижнем уровне $N_2 = 0,25/T$ и $N_6 = 0,75 \log 6/T = 1,95/T$, т. е. нагрузка элементов отличается почти в 8 раз.

Таким образом, иерархические структуры воплощают компромисс между противоречивыми стремлениями: с од-

ной стороны, обеспечить решение задачи возможно более простыми (дешевыми) средствами $C_k = C_{\min}$, а с другой стороны, обеспечить максимальную смысловую нагрузку (использование) на каждом из элементов $N_k = N_{\max}$.

Эти требования при неограниченном выборе элементов приводят, как следует из предыдущего, к иерархии, в которой на верхнем уровне находится самый простой элемент, а сложность (число состояний K , ветвей) элемента каждого последующего уровня определяется как два в степени общего числа ветвей предыдущего уровня (рис. 7-20, в): $\log K_{n+1} = K_n i$, где i — число элементов n -го уровня. Однако, если речь идет о человеческих системах, следует с учетом гипотезы Миллера ограничиться восемью ветвями в самом сложном элементе.

Определение затрат управленческого труда. До сих пор мы занимались оценкой только структурного смысла системы, характеризующего отношения частей и целого применительно к иерархическим структурам управления. Между тем управление связано с большим числом разного рода предварительных и вспомогательных операций, неизбежно сопутствующих решению основной проблемы. Так, например, почтальон, сортируя почту, т. е. осуществляя выбор из множества адресов своего участка, должен, как минимум, прочесть адреса на конверте, что связано с переработкой буквенной информации, большей чем структурная информация, равная согласно (7-52) числу элементов структуры, хотя системный смысл S обоих видов информации одинаков. Для буквенной или цифровой информации она равна числу букв или цифр, для словесной информации — числу произнесенных слов и т. д. Таким образом, реальные затраты управленческого труда, т. е. фактический смысл задачи S_{ϕ} , превосходят ее структурный смысл S во столько раз, во сколько раз фактически перерабатываемая для принятия решения информация J_{ϕ} превосходит структурную информацию J :

$$C_{\phi} = \sum_{i=1}^J J_{\phi i} H_i,$$

где H_i — сущность i -го элемента структуры; J_{ϕ} — информация того же элемента.

Точно так же фактическая смысловая мощность N_{ϕ} составляет $N_{\phi} = dC_{\phi}/dt$, а КПД управления $\eta = N/N_{\phi} = C/C_{\phi}$.

Фактические затраты умственного труда C_{Φ} оказывают существенное влияние на структуризацию задачи. Действительно, поскольку выбор в i -м элементе структуры на рис. 7-20 связан с усвоением (запоминанием) информации J_k (знаки, символы) в каждой ветви, то $J_{\Phi i} = K_i J_k$, где K_i — число ветвей i -го элемента.

Таким образом, ранее рассчитанные для элементов этих схем значения N увеличиваются в $J_{\Phi i}$ раз и могут превзойти допустимую для них смысловую мощность.

С другой стороны, если структуризация не изменяет полезный структурный смысл схемы, то она существенно уменьшает фактические затраты управленческого труда C_{Φ} . Действительно, если при тех же условиях для схемы a $C_{\Phi} = 8J_k \log 8 = 24J_k$, то для схемы b $C_{\Phi} = 3 \cdot 2J_k \log 2 = 6J_k$, т. е. в 4 раза меньше; для схемы $в$ $C_{\Phi} = 2J_k \log 2 + 4J_k \log 4 = 10J_k$; для схемы $г$ $C_{\Phi} = 4J_k \log 4 + 2J_k \log 2 = 10J_k$; для схемы $д$ в среднем $C_{\Phi} = 3J_k (-0,375 \log 0,375 - 0,25 \log 0,25 - 0,375 \log 0,375) + (0,375 \cdot 3J_k + 0,25 \cdot 2J_k + 0,375 \cdot 3J_k)(0,375 \log 3 + 0,25 \log 2 + 0,375 \log 3) = 8,6J_k$; для схемы $е$ в среднем $C_{\Phi} = 2J_k (-0,25 \log 0,25 - 0,75 \log 0,75) + (0,25 \times 2J_k + 0,75 \cdot 6J_k)(0,25 \log 2 + 0,75 \log 6) = 12,6J_k$.

Отсюда нетрудно сделать вывод, что любая структуризация уменьшает фактические затраты труда, что следует трактовать как еще одну причину структуризации. Кроме того, теперь наилучшей из всех представляется схема $б$, т. е. схема с самой большой из возможных в данном случае степеней структуризации, так что чем подробнее структуризация задачи, тем меньше затрат труда потребует ее решение и тем больше КПД управления, поскольку для схемы a $\eta = 0,125/J_k$, для схемы $б$ $\eta = 0,5/J_k$, для схемы $в$ $\eta = 0,3/J_k$ и т. д.

В заключение необходимо сказать о причинах структуризации, связанных с целями системы. С самых общих позиций цель есть отраженный системой объективный максимум функционала ее существования (жизнедеятельности). Объективный потому, что существуют объективные, не зависящие от степени их осознания условия, в наибольшей степени соответствующие потребностям жизнедеятельности системы, а отраженный потому, что система реально функционирует в соответствии со степенью осознания, понимания ее потребностей. Нередко объективный и отраженный максимумы практически совпадают, однако бывают случаи значительного расхождения между ними, как, например,

в случаях употребления людьми разного рода наркотиков, когда субъективно они ощущают улучшение своего состояния, но объективно наносят организму непоправимый вред.

Функционалом жизнедеятельности системы, максимум которого трактуется как ее цель, является вышеупомянутый смысл. При этом функционирование системы в целом имеет системный смысл C_c , совокупность элементов системы обладает собственным смыслом C_0 , а их разность образует взаимный смысл $C_{вз}$.

Как отмечалось, отрицательный взаимный смысл определяет устойчивость, стабильность системы. Чем больше $C_{вз}$ по абсолютному значению, тем больше стабильность системы, так что неструктуризованная, одноэлементная система,

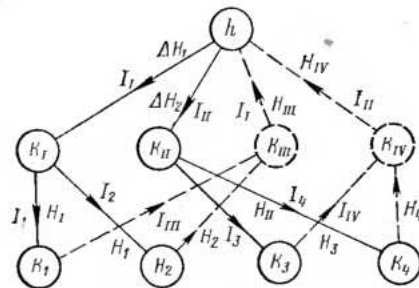


Рис. 7-21. Замкнутая иерархическая система

в которой системный смысл равен собственному смыслу, т. е. $C_{вз} = 0$, находится на границе устойчивости, а структуризованная система тем устойчивее, чем больше она структуризована, если, конечно, структуризация такова, что $C_{вз} < 0$. Таким образом, при одних и тех же целях $C_c = C_{с. макс}$ та из систем жизнедеятельности, которая больше структуризована. Это обстоятельство служит едва ли не основной причиной структуризации как естественных, так и социальных систем, хотя структуризация и снижает коэффициент использования элементов β .

Замкнутые системы. Любая иерархическая структура системы управления сопровождается организацией обратной связи в виде системы контроля, вместе с которой она образует всегда замкнутый контур.

На рис. 7-21 показан пример типичной древовидной иерархической структуры управления (сплошные линии) из трех уровней с разветвлениями в каждом уровне, рабо-

тающей совместно с аналогичной по структуре системой сбора информации (штриховые линии) и образующей с ней замкнутый контур управления и контроля, причем направления информационных токов показаны стрелками.

Такая структура примерно соответствует верхним уровням управления исследовательской организации, в которой директор имеет двух заместителей K_I и K_{II} , каждый из которых руководит деятельностью двух начальников отделов K_I и K_2 и соответственно K_3 и K_4 . Кроме того, в распоряжении директора имеются два консультанта K_{III} и K_{IV} , в обязанности которых входят контроль и оценка результатов деятельности отделов K_I и K_2 и соответственно K_3 и K_4 .

Символы K означают передаточные функции соответствующих руководителей при анализе структуры в переходном режиме (при переходе к новой теме исследований) или их коэффициенты усиления информационного напряжения (потенциалов) в установившемся режиме работы. Потенциалы всех уровней структуры в разомкнутом виде обычно легко установить. Так, если принять для общности, что отделы неравноценны и вносят неодинаковый вклад в общее дело и что вероятности выполнения нужной работы каждым отделом в отсутствие руководства составляют соответственно P_1, P_2, P_3 и P_4 , а с его участием они равны единице, то потенциалы начальников отделов (нижний уровень иерархии) составляют согласно (7-1) $H_1 = -\log P_1$, $H_2 = -\log P_2$ и т. д.

Потенциалы заместителей директора в этих условиях определяются совместной вероятностью достижения цели обоими подчиненными им отделами и составят $H_I = -\log P_1 P_2 = -\log P_1 - \log P_2 = H_1 + H_2$ и $H_{II} = -\log P_3 P_4 = -\log P_3 - \log P_4 = H_3 + H_4$. Наконец, потенциал директора, т. е. его ИДЛ, определится совместной вероятностью достижения цели обоими его заместителями, или, что то же самое, совместной вероятностью достижения цели всеми отделами: $h = -\log P_1 P_2 P_3 P_4 = H_I + H_{II} = H_1 + H_2 + H_3 + H_4$.

Таким образом, $K_I = H_I/h$; $K_{II} = H_{II}/h$; $K_1 = H_1/H_I$; $K_2 = H_2/H_I$; $K_3 = H_3/H_{II}$; $K_4 = H_4/H_{II}$.

В свою очередь, каждый из консультантов суммирует потенциалы подопечных отделов при подготовке доклада директору, снижая их критическими замечаниями в соответствии со своим коэффициентом усиления, так что $H_{III} = (H_1 + H_2) K_{III}$, а $H_{IV} = (H_3 + H_4) K_{IV}$, где K_{III} и K_{IV} —

это отношения потенциалов отчетов отделов к их потенциалам после критического анализа.

Директор же управляет своими заместителями, руководствуясь критическим анализом их суммарных достижений, т. е. беспокоит их тем меньше, чем успешнее они работают: $\Delta H_1 = h - H_{III}$; $\Delta H_2 = h - H_{IV}$, в чем и проявляется действие отрицательной обратной связи системы контроля.

Предположим теперь, что в течение всего рабочего времени T_p директор в среднем занят со своими заместителями соответственно t_I и t_{II} , а со своими консультантами t_{III} и t_{IV} , причем $T_p > t_I + t_{II} + t_{III} + t_{IV}$ (если у директора есть еще и другие дела) и $T_p < t_I + t_{II} + t_{III} + t_{IV}$ (если он иногда беседует сразу с несколькими подчиненными).

Если время общения по каждому конкретному поводу (информационное сопротивление) с каждым из заместителей и консультантов составляет соответственно $\tau_I, \tau_{II}, \tau_{III}$ и τ_{IV} , то информационные токи составляют $I_I = \Delta H / (\tau_I + \tau_{III})$; $I_{II} = \Delta H / (\tau_{II} + \tau_{IV})$.

Точно по такой же логике токи между заместителями и начальниками отделов составляют

$$I_1 = H_1 t_1 / (T_p \tau_1); \quad I_2 = H_2 t_2 / (T_p \tau_2);$$

$$I_3 = H_3 t_3 / (T_p \tau_3); \quad I_4 = H_4 t_4 / (T_p \tau_4);$$

токи между отделами и консультантами:

$$I_{III} = (H_1 + H_2) t'_{III} / (T_p \tau'_{III}); \quad I_{IV} = (H_3 + H_4) t'_{IV} / (T_p \tau'_{IV}),$$

где τ'_{III} и τ'_{IV} — время для изучения отчетности подопечных отделов по конкретному акту управления каждым из консультантов.

По этим токам и потенциалам можно определить смысловую мощность $N = \Delta H I$ работников всех уровней, сопоставив ее либо с предельно допустимой для человека, либо со средней принятой в данной отрасли.

Таким образом, для информационного анализа незамкнутой иерархической структуры необходимо знание только вероятностей P_1, P_2, P_3, P_4 и бюджета времени работников всех рангов. Однако такой анализ отдельных ветвей и узлов иерархии будет неполным, а для замкнутой системы и неверным, поскольку не дает возможности сравнивать как целое различные варианты структур управления. С этой точки зрения наиболее удобны передаточные функции (в установившемся режиме — коэффициенты усиления напряжения) систем, которые получаются в результате решения системы

полученных выше уравнений, связывающих ИДЛ и потенциалы уровней, или исходя из структуры на рис. 7-21, где K_1 и K_2 соединены параллельно между собой и последовательно с K_{III} , т. е. их эквивалентная передаточная функция равна $K_I(K_1 + K_2)$. Точно так же соединены K_{II} , K_3 и K_4 , поэтому их передаточная функция равна $K_{II}(K_3 + K_4)$. В свою очередь, каждая из этих цепочек охвачена обратной связью с коэффициентом соответственно K_{III} и K_{IV} , т. е. согласно (7-51) они имеют эквивалентные передаточные функции

$$\frac{K_I(K_1 + K_2)}{1 + K_I(K_1 + K_2)K_{III}}; \quad \frac{K_{II}(K_3 + K_4)}{1 + K_{II}(K_3 + K_4)K_{IV}}.$$

Поскольку оба этих замкнутых контура параллельны друг другу, то передаточная функция всей схемы, представляющая собой в установившемся режиме отношение истинного суммарного потенциала всех отделов замкнутой системы H_Σ к ИДЛ источника, составляет

$$\frac{H_\Sigma}{h} = \frac{K_I(K_1 + K_2)}{1 + K_I(K_1 + K_2)K_{III}} + \frac{K_{II}(K_3 + K_4)}{1 + K_{II}(K_3 + K_4)K_{IV}}.$$

Из этой передаточной функции следует, что $H_\Sigma = h = H_1 + H_2 + H_3 = H_4$, только если система работает без обратной связи, когда $K_{III} = K_{IV} = 0$; во всех же других случаях $H_\Sigma < h$. Так, если принять $P_1 = P_2 = P_3 = P_4$ и $K_{III} = K_{IV} = K_{o.c.}$, получим $K_I = K_{II} = K_1 = K_2 = K_3 = K_4 = 0,5$, откуда $H_\Sigma/h = 1/(1 + 0,5K_{o.c.})$.

Если $K_{o.c.} = 1$, т. е. если консультанты правильно оценивают деятельность отделов, то $H_\Sigma = 2h/3$; если же они перехваливают отделы ($K_{o.c.} > 1$), то $H_\Sigma < 2h/3$. При критическом отношении к работе отделов, когда $K_{o.c.} < 1$, имеем $H_\Sigma > 2h/3$. Наконец, в отсутствие обратной связи ($K_{o.c.} = 0$), что равносильно для директора отсутствию оценки его управления, $H_\Sigma = h$.

Отметим еще, что в случае положительной обратной связи, когда консультанты извращают результаты управления ($K_{o.c.} < 0$), можно получить $H_\Sigma > h$, а при $K_{o.c.} = -2$, т. е. при умышленно противоположном толковании результатов, даже $H_\Sigma \rightarrow \infty$, что сразу перегрузит работников всех уровней и развалит систему.

Как уже отмечалось, рассчитанные выше потенциалы и токи в общем случае пригодны только для определения коэффициентов усиления, но не смысловой нагрузки ветвей и узлов в замкнутой системе, поскольку при замыкании

системы обратной связью эти напряжения и токи изменяются в соответствии с коэффициентом усиления (передаточной функцией) всей замкнутой системы. Для определения истинных значений напряжений и токов замкнутой системы нужно замкнуть систему, исходя из того, что передаточные функции отдельных звеньев и ИДЛ источника при замыкании не изменяются. Так, обратившись к составляющим систему передаточным функциям отдельных замкнутых контуров

$$\frac{H_1 + H_2}{h} = \frac{K_I(K_1 + K_2)}{1 + K_I(K_1 + K_2)K_{III}};$$

$$\frac{H_3 + H_4}{h} = \frac{K_{II}(K_3 + K_4)}{1 + K_{II}(K_3 + K_4)K_{IV}},$$

нетрудно получить по свойству пропорций

$$\frac{\Delta H_1}{h} = \frac{h - (H_1 + H_2)}{h} = \frac{1 + K_I(K_1 + K_2)(K_{III} - 1)}{1 + K_I(K_1 + K_2)K_{III}};$$

$$\frac{\Delta H_2}{h} = \frac{h - (H_3 + H_4)}{h} = \frac{1 + K_{II}(K_3 + K_4)(K_{IV} - 1)}{1 + K_{II}(K_3 + K_4)K_{IV}},$$

откуда найдутся ΔH_1 и ΔH_2 . Тогда: $H_1 = K_I \Delta H_1$; $H_{II} = K_{II} \Delta H_2$; $H_1 = K_I H_1$; $H_2 = K_2 H_1$; $H_3 = K_3 H_{II}$; $H_4 = K_4 H_{II}$. Подставив эти напряжения в формулы для токов, нетрудно определить значения последних в замкнутой системе, а вместе с ними и истинную смысловую мощность элементов в замкнутой системе.

7-5. УПРАВЛЕНИЕ СИСТЕМАМИ ПОДВИЖНЫХ ОБЪЕКТОВ

Существует класс больших систем управления положением или движением множества подвижных объектов в том или ином пространстве. К классу таких систем относятся системы управления уличным движением в масштабах города, движением самолетов в воздушном пространстве вблизи аэропортов, воздушными боями и встречными боями механизированных наземных частей, размещением судов в гаванях и элементов в конструкциях, производительных сил в стране, шахматных фигур на доске и т. п.

Помимо общих для всех БСУ проблем переработки в ограниченное время больших объемов информации, этот класс систем обладает еще тем свойством, что максимально успешное управление в отдельных частях пространства может оказаться неуспешным для всей системы в целом. Такого рода ситуации хорошо известны шахматистам, когда стратегия сохранения фигур оказывается менее эффективной, нежели стратегия, связанная с жертвами фигур с целью получения решающего перевеса,

Кроме того, при быстрых перемещениях объектов в больших пространствах существенным оказывается запаздывание обмена информацией между отдельными частями системы, так что управление вынуждено прогнозировать текущее состояние тех или иных частей системы по их предшествующему поведению и состоянию других частей [17].

Наконец, с формальной точки зрения особенность систем этого класса состоит еще в том, что они обычно описываются дифференциальными уравнениями в частных производных, неудобных для ЦВМ.

Особенно остро стоят проблемы математического описания в человеко-машинных системах, когда, помимо законов физики, приходится учитывать еще логику поведения операторов.

Ниже демонстрируются некоторые аспекты управления в системах такого типа на примере воздушного боя большого числа самолетов.

Позиция, строй, статика. Рассмотрим вначале только такие ситуации, которые либо вовсе не изменяются на интересующем нас отрезке времени, либо изменяются настолько медленно, что нет необходимости учитывать относительные скорости объектов, а достаточно ограничиться изучением их позиции, строя, т. е. взаимного расположения. Это ограничение вовсе не касается скорости движения строя (или системы) в целом, которая при этом может быть сколь угодно велика, если только скорость движения сама по себе не определяет строя, о чем речь пойдет несколько ниже.

Для описания статистики системы воспользуемся одним из основных положений материалистической диалектики о познаваемости мира, согласно которому окружающий нас материальный мир адекватно отражается, «фотографируется, копируется» нашими органами чувств. Назовем это положение законом отражения и запишем его в форме

$$J = R_k M, \quad (7-57)$$

где M — количество интересующей нас материи объекта, которое мы для краткости будем именовать просто материей; J — актуальная информация, которую мы в данных условиях способны получить о материи; R_k — информационная проницаемость, характеризующая условия отражения, состояние технического оснащения и т. п.

Здесь речь идет не о любых информации и материи, а лишь об актуальных с точки зрения целей управления. Неактуальные материи и информацию мы просто игнорируем в процессе управления, создавая для них $R_k = 0$.

В форме (7-57) материя M численно равна той актуальной информации, которая могла бы быть получена об интересующих нас свойствах объекта в идеальных условиях, т. е. при неограниченных технических возможностях (вещь в себе), а информация J — это лишь та часть M , которая доступна восприятию в реальных условиях (вещь для нас). Таким образом, в реальных условиях обычно $J \leq M$, а $0 \leq R_k \leq 1$. Однако, если противник прибегает не только к маскировке, уменьшающей J по сравнению с M , но и к созданию ложных позиций или ложных целей на экранах локаторов, то может оказаться $J > M$, а $R_k > 1$.

Информацию и материю мы будем измерять в битах, причем, как показано ранее, для дискретных объектов они просто равны числу соответственно обнаруженных и реально имеющихся объектов. В случаях, обычно имеющих место, когда информация поступает от несколь-

ких приборов и по нескольким каналам, вместо (7-57) имеем

$$M = \sum_{k=1}^n J_k / R_k,$$

где n — число каналов;

или в пределе при непрерывном обследовании объекта со всех сторон:

$$M = \oint_S (O / R_k) dS, \quad (7-58)$$

где S — произвольная замкнутая вокруг изучаемого объекта поверхность; O — плотность потока актуальной информации (отражения), т. е. информация, приходящаяся на единицу площади замкнутой поверхности в каждой ее точке, причем $O = dJ/dS$.

Наконец, в дифференциальной (локальной) форме применительно к каждой точке пространства закон отражения принимает вид:

$$\text{div } O = R_k \rho, \quad (7-59)$$

где ρ — объемная плотность материи в данной точке пространства, причем $\rho = \partial M / \partial V$; V — объем пространства.

Закон отражения позволяет по произвольно заданному в пространстве распределению материи (объектов) определить плотность O потока информации о них в каждой точке пространства, что представляет самостоятельный интерес при необходимости обеспечить наилучшие условия наблюдения за строем для руководителя полетов или, наоборот, обеспечить наихудшие условия для наблюдателей противника.

Так, например, из (7-58) следует, что для наблюдения со стороны безразлично, собрать ли все M самолетов в плотную группу или распределить их внутри большой сферы радиусом r_0 , симметрично относительно ее центра, поскольку ввиду сферической симметрии для всех $r \geq r_0$ в обоих случаях

$$M = (O / R_k) \oint_S dS = 4\pi r^2 O / R_k, \text{ или } O = R_k M / (4\pi r^2), \quad (7-60)$$

т. е. модуль плотности потока информации вне сферы распределения самолетов обратен квадрату расстояния от центра сферы до наблюдателя (если, конечно, R_k не зависит от r). При этом, поскольку в любой сфере, проведенной внутри сферы сплошного распределения самолетов ($r \leq r_0$) из их общего центра, $M_r = M r^3 / r_0^3$, то согласно (7-60) для наблюдателя, расположенного в любой точке внутри зоны распределения, плотность потока информации составит $O = M R_k r / (4\pi r_0^3)$, т. е. она увеличивается от центра к периферии, так что с учетом (7-60) наилучшие условия наблюдения имеют место на границе области распределения самолетов, т. е. при $r = r_0$.

Подчеркнем еще раз, что речь идет лишь об актуальной, т. е. управляющей информации, которая способна указать направление тех или иных целенаправленных действий. По этой причине в центре распределения объектов наблюдения $O = 0$, так как ввиду центральной симметрии здесь совершенно равнозначны действия, предпринятые в любом направлении, т. е. невозможно указать привилегированное поведение.

Мы здесь, по существу, решаем оптимизационную задачу $O \rightarrow \text{extr}$, что с учетом соответствующих ограничений описывает процедуру управ-

ления наблюдателем группой самолетов. Необходимо лишь добавить, что при n компактных групп объектов на расстояниях, значительно превышающих линейные размеры этих групп, вместо (7-60), естественно, имеем

$$\mathbf{O} = \sum_{i=1}^n [R_k M_i \mathbf{r}_i / (4\pi r_i^3)], \quad (7-61)$$

где M_i и r_i — соответственно число объектов в i -й группе и расстояние до нее от наблюдателя.

Если же положение наблюдателя фиксированно, то приходим к задаче управления положением (строем) объектов по критерию наилучшей (или наихудшей для противника) наблюдаемости в заданной точке пространства

$$\mathbf{O} \Rightarrow \text{extr} \quad (7-62)$$

с учетом (7-61) или в произвольных случаях (7-59) и имеющихся ограничений. Например, если имеет место (7-61), а ближайшее расстояние до наблюдателя ограничено $r_{\min}$, то, поскольку в соответствии с (7-62) $r_k \rightarrow 0$, приходится принять все $r_i = r_{\min}$, т. е. расположить наблюдаемые группы объектов в одной точке на сферической поверхности с центром в месте расположения наблюдателя, как можно ближе к нему. Если речь идет о группах разнотипных объектов, наблюдателя (диспетчера, оператора) обычно интересует определенный набор их материальных свойств (размеры, вооружение и другие тактико-технические характеристики), наблюдаемых в различной степени и к тому же нередко непрерывных (масса, габариты и пр.), т. е. не дискретных. В этом случае приходится говорить об объекте как о векторе материи $\mathbf{M}$ ($M_1, M_2, \dots, M_k, \dots, M_m$), компоненты M_k которого представляют собой обычные (скалярные) материи (число членов экипажа, число двигателей, число пушек и т. д.), если они дискретны; либо, если соответствующие свойства непрерывны, $M_k = A_k / \Delta A_k$, где A_k — непрерывная физическая величина (габарит, масса, запас горючего, стоймость и т. п.), а ΔA_k — принятый в данных условиях минимальный имеющий смысл размер («квант») соответствующей физической величины (метр, тонна, тысяча рублей и пр.).

Поскольку на расстояниях от наблюдателя, существенно больших линейных размеров объекта, вся его материя практически сосредото-

чена в одной точке, то вместо (7-57) имеем $\oint \mathbf{O} dS = \sum_{k=1}^m R_k M_k$ и $\text{div } \mathbf{O} =$

$$= \sum_{k=1}^m R_k \rho_k, \quad \text{а вместо (7-60), } \mathbf{O} = \sum_{k=1}^m R_k M_k / (4\pi r^2) \quad \text{и } \mathbf{O} =$$

$$= \sum_{i=1}^n \left[\mathbf{r}_i / (4\pi r_i^3) \sum_{k=1}^m R_{ki} M_{ki} \right], \quad \text{так что при взаимнезависимости } M_k$$

процедура управления по-прежнему сводится к (7-62).

Однако для нескольких или большого числа наблюдателей рассматриваемый аппарат управления недостаточен, как недостаточен он и для воздушного боя, где каждый из его участников является, с одной стороны, наблюдателем, а с другой стороны, объектом наблюдения для остальных участников. В этом случае каждая точка пространства характеризуется плотностью потока актуальной информации $\mathbf{O}$ и сосредоточенной в этой точке материей M , способной реагировать

на поток информации. Перемножив эти величины, получим вектор плотности смыслового потока δ для этой точки, т. е. знакомый читателям по информационному анализу иерархий смысл управления в расчете на единицу поверхности, нормальной потоку:

$$\delta = \mathbf{O}M = dC/dS. \quad (7-63)$$

Учитывая, что вектор δ указывает в каждой точке пространства направление и размер изменений в системе, желательных с точки зрения достижения цели управления, в дальнейшем будем называть его вектором управления. Таким образом, вектор управления дает программу изменения ситуации в каждой точке системы в направлении к поставленной цели, которая характеризуется $\delta = 0$, поскольку при достижении цели системой любые изменения ее состояния нежелательны, ибо уводят от цели.

Управление согласно (7-63) требует изменения ситуации в опасных точках (где велика $\mathbf{O}$) не только путем воздействия на противника, но и простым удалением своих объектов из этих точек, после чего в них $\delta = 0$ за счет $M = 0$. Такого рода ситуация имеет место, например, когда один самолет M_1 ведет бой с большим числом M самолетов противника, движение которых образует близкий к шару клубок, вне которого, как было показано, $\mathbf{O} = R_k M / (4\pi r^2)$, где r — расстояние от центра, а внутри $\mathbf{O} = R_k M r / (4\pi r^3)$, так что оказавшемуся в одиночестве самолету приходится либо спасаться бегством в направлении $r = \infty$, где $\mathbf{O} = \delta = 0$, либо рискнуть вступить в бой в центре клубка, где $r = 0$ и где он почти неуязвим, так как противник рискует при стрельбе поразить свои самолеты. В последнем случае, однако, противник будет прилагать усилия к тому, чтобы, изменяя свое расположение, удалить этот самолет из середины своего строя на периферию, где он сразу будет уничтожен, так как там $\delta = \delta_{\max} = R_k M M_1 / (4\pi r^3)$.

На этом примере хорошо видно, что поскольку противники преследуют противоположные цели, то с математической точки зрения один из них стремится минимизировать δ в местах своего расположения, а другой, напротив, добивается $\delta = \delta_{\max}$ в тех же местах, параллельно заботясь о том, чтобы в местах своего расположения сделать $\delta = 0$.

Стратегия (7-63), однако, чревата возможностью выбрать лишь относительно хорошую, но не самую лучшую позицию. Действительно, может возникнуть относительно благоприятная ситуация, когда в определенной точке $\delta = 0$, т. е. малые отклонения от нее ухудшают ситуацию, однако большие перемещения могут привести в другую точку с $\delta = 0$, но с более благоприятными условиями. Поскольку в обеих точках вектор управления равен нулю, то отличить их можно, лишь обращаясь к сущности (смыслу) ситуации в них.

Зададимся неким «квантом» длины R_0 , характерным для изучаемых отношений между объектами, и помножим числитель и знаменатель (7-63) на вектор R_0 , совпадающий по направлению с dS , приняв во внимание также, что $dS = R_0^2$ и согласно разделу об анализе иерархий $C = MH$:

$$\delta = (R_0 M / dS) (dH / R_0) = - (M / R_0) \text{grad } H.$$

Сравнивая полученное выражение с (7-63), имеем:

$$R_0 \mathbf{O} = - \text{grad } H, \quad (7-64)$$

что позволяет перейти от поля вектора управления δ к скалярному полю сущности H или смысла C . Подставляя (7-64) в (7-59), получим

уравнение Пуассона, позволяющее посредством $H = -\log P$ определить сущность, т. е. вероятность недостижения цели управления, в каждой точке пространства по заданному распределению ρ материи:

$$\operatorname{div} \operatorname{grad} H = \Delta H = -R_k R_0 \rho = -R\rho, \quad (7-65)$$

где $R_k R_0 = R$ — информационная проницаемость пространства; R_k — безразмерная относительная информационная проницаемость; R_0 — абсолютная информационная проницаемость, измеряемая в метрах.

Решив уравнение Пуассона (7-65), можно перейти к объемной плотности c смысла ситуации в каждой точке пространства, которая учитывает не только вероятность достижения цели управления, но и число объектов в единице объема, для которых суть ситуации одна и та же:

$$c = \rho H, \quad (7-66)$$

так что управление требует $c \Rightarrow c_{\max}$, в то время как противник в условиях конфликта добивается $c \Rightarrow c_{\min}$.

В отличие от (7-63) управление по (7-66) не может ограничиться частичным благополучием «в овраге» $\delta = 0$, но требует достижения глобального $c_{\max}$. Это не изменяет техники управления, ибо все перемещения объектов по-прежнему должны совершаться в соответствии с δ , но не в точки с $\delta = 0$, а в точки с максимально возможным в условиях ограничений значением c , т. е. в точки с максимально возможной в данных условиях вероятностью достижения цели для возможно большего числа объектов. Кроме того, относительно простой критерий (7-66) предполагает, что если в каждой точке пространства ситуация максимально благополучна (с учетом ограничений), то максимально благополучна и ситуация во всем пространстве, т. е. для системы в целом. Однако это предположение далеко не всегда оправдывается, ибо, например, стратегия сознательных жертв в одних точках во имя решающего выигрыша в целом этим критерием не предусматривается.

Действительно, универсальным критерием управления является лишь смысл C всей ситуации в целом

$$C = \int_V c dV, \quad (7-67)$$

максимизация которого обеспечивает предельную в данных условиях вероятность достижения глобальной цели управления системой.

С учетом частного решения уравнения Пуассона вместо (7-67) можно также записать

$$C = \frac{1}{4\pi} \iint_V \frac{R\rho_1 \rho_2}{r_{12}} dV_1 dV_2,$$

причем, как и во всякой системе, смысл может быть собственным C_0 , системным C_c и взаимным $C_{вз}$. Например, для двух компактных групп объектов M_1 и M_2 имеем соответственно $O_1 = R_k M_1 / (4\pi r^2)$ и $O_2 = R_k M_2 / (4\pi r^2)$ или с учетом (7-65): $H_1 = R M_1 / (4\pi r)$ и $H_2 = R M_2 / (4\pi r)$, откуда собственные смыслы групп и системы в целом составят

$$C_1 = M_1 H_1, \quad C_2 = M_2 H_2 \quad \text{и} \quad C_0 = C_1 + C_2 = R (r_2 M_1^2 + r_1 M_2^2) / (4\pi r_1 r_2),$$

где r_1 и r_2 — радиусы групп, а взаимные смыслы: $C_{вз} = C_{12} = C_{21} = M_1 H_2 = M_2 H_1 = R M_1 M_2 / (4\pi r_{12})$, где r_{12} — расстояние между центрами групп.

Таким образом, системный смысл составит $C_c = C_0 + C_{вз} = R (r_2 M_1^2 + r_1 M_2^2 + r_{12} M_1 M_2) / (4\pi r_1 r_2 r_{12})$, а целостность системы $\alpha = -C_{вз}/C_0 = -M_1 M_2 r_1 r_2 / (r_{12} M_1^2 + r_{12} M_2^2)$ отрицательна, если обе группы принадлежат одной из противоборствующих сторон ($\operatorname{sign} M_1 = \operatorname{sign} M_2$), и положительна, если группы принадлежат противоположным сторонам ($\operatorname{sign} M_1 = -\operatorname{sign} M_2$). В частности, если группы одинаковы, т. е. $M_1 = M_2$ и $r_1 = r_2 = r_0$, целостность системы $\alpha = -r_0 / (2r_{12})$ по абсолютному значению тем больше, чем ближе группы между собой и чем они больше.

Во всех этих соотношениях информационная проницаемость R определяется логикой отношений между объектами, а также условиями наблюдения. Однако она может быть определена и экспериментально без проникновения в суть этих отношений. Действительно, с позиций воздушного боя, например, нас интересует вероятность P поражения одним самолетом другого, т. е. для компактной группы из M объектов $H = -\log(1 - P) = RM / (4\pi r)$, что после экспериментальных стрельб по мишеням на расстоянии r от группы позволяет рассчитать R по вероятности поражения мишеней.

Абсолютная проницаемость R_0 может быть также определена из соотношения между смыслами наблюдения и управления $\delta = R_0 c$, которое следует из (7-63).

Управление стационарным движением объектов. Рассмотрим теперь такие стационарные ситуации, в которых не меняется относительное расположение объектов, но скорости движения объектов играют определенную роль, поскольку запас кинетической энергии обеспечивает свободу маневра.

Для описания таких ситуаций закон отражения материи (7-57) необходимо дополнить еще законом сохранения материи

$$\oint_S \mathbf{j}_m d\mathbf{S} = -dM/dt \quad (7-68)$$

или в дифференциальной форме

$$\operatorname{div} \mathbf{j}_m = -dp/dt, \quad (7-69)$$

где $\mathbf{j}_m$ — плотность материального тока, т. е. материя, переносимая в единицу времени сквозь единицу поверхности.

Соотношение (7-68) подтверждается всем историческим опытом человечества и получается при предельном переходе из соотношения $\Sigma I_k = -dM/dt$, которое означает, что суммарный ток материи сквозь замкнутую поверхность равен убыли материи внутри этой поверхности.

Однако, поскольку нас пока интересуют только стационарные ситуации, то в соотношениях (7-68) $dM/dt = 0$. При этом, поскольку всегда $\operatorname{div} \operatorname{rot} \mathbf{I} = 0$, из (7-68) следует закон отражения движения:

$$a_k \operatorname{rot} \mathbf{I} = \mathbf{j}_m \quad \text{или} \quad \oint_l a_k \mathbf{I} dl = I_m, \quad (7-70)$$

где I_m — материальный ток, т. е. материя, перемещающаяся сквозь произвольный замкнутый контур длиной l в единицу времени; $\mathbf{I}$ — вектор индукции, т. е. информационный (отраженный контуром) ток, приходящийся на единицу длины контура; a_k — относительная целевая проницаемость среды, характеризующая отношение между отражаемым и отраженным движением (токами),

Закон отражения движения позволяет по заданному в каждой точке пространства движению (плотности тока материи) определить условия наблюдения (вектор индукции) во всех точках. Поскольку, однако, смысл наблюдения определяется не только индукцией в той или иной точке, но и материей, способной воспринимать его в той же точке, то имеем для уже знакомой по анализу структур смысловой мощности N на единицу длины произвольного контура

$$\xi = \mathbf{I}M = dN/dl, \quad (7-71)$$

причем управление наблюдением сводится к $\xi \Rightarrow \xi_{\text{макс}}$.

Так, для прямолинейного движения достаточно длинной колонны, сквозь неподвижное поперечное сечение которой в каждую секунду пролетает I_M самолетов, получим на окружности радиуса r , охватывающей ось колонны в плоскости, перпендикулярной оси: $I_M = 2\pi r_k I$ и $\xi = I_M M / (2\pi r_k)$, где M — число наблюдателей на окружности.

Из полученных выражений следует очевидный вывод, что наблюдатели должны располагаться возможно ближе к колонне с учетом возможных ограничений (опасности). Если наблюдатели окажутся внутри колонны на расстоянии r от оси, то внутри концентрической окружности, на которой они находятся, окажется ток $I_M r^2 / r_0^2$, где r_0 — радиус колонны, так что $\xi = I_M M r / (2\pi r_k^2 a_k)$. В этом случае наилучшие условия для них создадутся на периферии колонны при $r = r_0$.

Соотношение (7-71) пригодно лишь для оптимизации наблюдения, но не для управления движением во имя достижения определенной цели, которое должно характеризоваться самостоятельной абсолютной целевой проницаемостью a_0 .

Действительно, умножив числитель и знаменатель (7-71) на $\mathbf{V} \times \mathbf{R}_0$, где $\mathbf{V}$ — скорость движения материи, получим

$$\xi = (\mathbf{v} \times \mathbf{R}_0) dN / (v R_0 dl) = (\mathbf{v} \times \mathbf{R}_0) c = \mathbf{v} \times \delta. \quad (7-72)$$

С другой стороны, плотность смысла в каждой точке можно представить в виде произведения плотности тока в этой точке и векторного потенциала $\mathbf{A}$, представляющего собой сущность материи, движущейся со скоростью l м/с, так что

$$c = j_M \mathbf{A} \text{ и } C = \mathbf{A} I, \quad (7-73)$$

где l — длина элемента тока I .

Подставив (7-73) в (7-72), можно получить:

$$\mathbf{I} = a_0 \text{ rot } \mathbf{A}. \quad (7-74)$$

Наконец, (7-74) с учетом (7-70) дает

$$j_M = a \text{ rot rot } \mathbf{A} = -a \Delta \mathbf{A}, \quad (7-75)$$

где $a_h a_0 = a$ — целевая проницаемость среды.

Полученное соотношение (7-75) позволяет по заданному распределению движения в пространстве, т. е. по плотности материальных токов, определить векторный потенциал $\mathbf{A}$, а затем посредством (7-73) — суть движения и плотность смысла движения в каждой точке. При этом простейшая стратегия управления движением сводится к $c \Rightarrow \xi_{\text{макс}}$ в каждой точке пространства, что чревато потерей наилучших для системы как целого вариантов ее движения. Для организации оптимального движения системы наилучшая стратегия $C \Rightarrow C_{\text{макс}}$ обеспечивается соотношением (7-67) с подстановкой в него плотности смысла движения (7-73), что для частного решения (7-75) приводит к форме

$$C = \frac{1}{4\pi} \iint_V \frac{j_1 j_2}{ar_{12}} dV_1 dV_2, \text{ причем смысл движения, так же как раньше,}$$

может быть собственным, системным и взаимным. Например, при параллельном движении двух авиационных колонн I_1 и I_2 со скоростями V_1 и V_2 на расстоянии r друг от друга имеем соответственно $I_1 = I_1 / (2\pi r a_k)$ и $I_2 = I_2 / (2\pi r a_k)$ или с учетом (7-74) $A_1 = I_1 l_1 / (2\pi r a)$ и $A_2 = I_2 l_2 / (2\pi r a)$, чему согласно (7-73) соответствует собственный смысл движения системы $C_0 = C_1 + C_2 = (I_1 l_1)^2 / (2\pi r_1 a) + (I_2 l_2)^2 / (2\pi r_2 a)$, где r_1 и r_2 — радиусы колонн.

Взаимный смысл движения колонн $C_{\text{вз}} = C_{12} = C_{21} = I_1 I_2 l_1 l_2 / (2\pi r_{12} a)$.

Таким образом, системный смысл движения составит $C_c = C_0 + C_{\text{вз}} = (I_1^2 l_1^2 r_{12} + I_1 I_2 l_1 l_2 r_1 r_2 + I_2^2 l_2^2 r_{12}) / (2\pi r_1 r_2 r_{12} a)$, а кинетическая целостность $\alpha = -C_{\text{вз}} / C_0 = I_1 I_2 l_1 l_2 r_1 r_2 / (I_1^2 l_1^2 r_{12} + I_2^2 l_2^2 r_{12})$, что для одинаковых колонн, т. е. при $I_1 = I_2 = I$, $l_1 = l_2 = l$, $r_1 = r_2 = r_0$, дает $\alpha = -r_0 / (2r_{12})$. Конечной целью такого управления является достижение $C_{\text{макс}}$, чему соответствует $\text{grad } C = 0$. Однако, помимо цели, нужно указать еще пути ее достижения. Такое указание дает вектор логической связи (логики управления) $\mathbf{J} = -\text{grad } C$, или с учетом (7-67)

$$\mathbf{J} = - \int_V \text{grad } c dV = \int_V l dV, \quad (7-76)$$

где l — объемная плотность логики управления.

Так, для параллельных колонн $\mathbf{J} = -\text{grad } (C_{\text{вз}} = I_1 I_2 l_1 l_2 / (2\pi r a^2))$, т. е. логика боя требует сближения противоборствующих колонн ($\text{sign } l_1 = -\text{sign } l_2$) и разведения колонн одной стороны ($\text{sign } l_1 = \text{sign } l_2$).

Разумеется, градиент смысла может браться по любому параметру, а не только по r , т. е. по расстоянию между объектами. В частности, можно оптимизировать смысл по скоростям (токам) движения объектов, размерам l_k , r_k колонн и т. д. При этом характеристическая константа a может быть вычислена на основании экспериментально определенных информационной проницаемости R и скорости v носителей взаимодействия объектов (снарядов, ракет и др.):

$$v^2 = Ra. \quad (7-77)$$

До сих пор мы рассматривали по отдельности друг от друга взаимодействие неподвижных и движущихся объектов. В реальных условиях стационарно-движущиеся объекты в процессе движения взаимодействуют и как неподвижные объекты. По этой причине реальное управление должно исходить из совокупности этих факторов, т. е. из соотношения $C = \int (\rho H + \mathbf{A} j_M) dV$, которое легко может быть

приведено к статической форме $C = \int_V (\rho H + \rho_v H_v) dV$, если обозначить $\rho_v = j_M / v$ и $H_v = Av$, где v — скорость движения материи в каждой точке, совпадающая по направлению с j_M .

Управление боем. Особенностью воздушного боя является стремительность и быстротечность, вследствие чего ситуация в воздухе быстро меняется и не может считаться стационарной. С формальной точки зрения это приводит лишь к необходимости учета в (7-68) $\partial M / \partial t \neq 0$. Однако все предыдущие соотношения исходят из $\partial M / \partial t = 0$, так что, на первый взгляд, потребуется выработка другого аппарата описания

движения. Между тем, если продифференцировать по времени (7-57) и ввести понятие плотности тока информации (отражения) $\partial \mathbf{O} / \partial t = R_k \mathbf{j}_0$, то получим $\operatorname{div} \mathbf{j}_0 = \partial \rho / \partial t$. Сравнивая последнее соотношение с (7-69), получаем:

$$\operatorname{div} (\mathbf{j}_m + \mathbf{j}_0) = 0, \text{ или } \operatorname{div} \mathbf{j} = 0, \quad (7-78)$$

где $\mathbf{j} = \mathbf{j}_m + \mathbf{j}_0$, т. е. полный ток равен сумме материального и информационного токов.

Соотношение (7-78) представляет собой закон полного тока, т. е. закон сохранения материи и энергии. Поскольку (7-78) по форме аналогично (7-68), для которого справедливы полученные ранее соотношения, начиная с (7-70) и далее, то они справедливы и для (7-78), если в них повсюду заменить плотность материального тока $\mathbf{j}_m$ плотностью полного тока $\mathbf{j}$.

Кроме того, поскольку теперь $\partial A / \partial t \neq 0$, то в (7-64) градиент сущности H складывается не только из логики положения $R_0 \mathbf{O}$, но и из логики тенденций изменения $\partial A / \partial t$, так что $R_0 \mathbf{O} + \partial A / \partial t = -\operatorname{grad} H$.

Таким образом, смысл управления формально описывается тем же соотношением

$$C = \int_V (\rho H + A \mathbf{j}) dV. \quad (7-79)$$

в котором $\mathbf{j} = \mathbf{j}_m + \mathbf{j}_0$, $\mathbf{j}_m = \rho \mathbf{v}$, $R_k \mathbf{j}_0 = \partial \mathbf{O} / \partial t$, $\operatorname{div} \mathbf{O} = R_k \rho$, $R_0 \mathbf{O} + \partial A / \partial t = -\operatorname{grad} H$, $a_k \operatorname{rot} \mathbf{H} = \mathbf{j}$, $\mathbf{H} = a_0 \operatorname{rot} \mathbf{A}$, а стратегия управления обеспечивается $C \rightarrow C_{\max}$ посредством (7-76) с учетом возможных ограничений.

Рассмотренный в данной главе информационный подход является новым направлением исследований в теории БСУ, позволяющим с единой информационной точки зрения описать процессы функционирования больших систем различной физической природы и назначения.

ГЛАВА ВОСЬМАЯ

СТРУКТУРНЫЙ СИНТЕЗ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ

8-1. АГРЕГАТИЗАЦИЯ — ОСНОВА ПОСТРОЕНИЯ СТРУКТУРЫ СОВРЕМЕННЫХ БСУ

В настоящее время при построении структуры БСУ широко используются агрегатированные комплексы унифицированных элементов. Первые системы автоматизированного управления проектировались для решения определенной задачи или ограниченного класса задач, и структура систем была жесткой, с трудом поддающейся даже небольшим изменениям. Однако со временем стало ясно, что путь индивидуального проектирования систем малоперспективен

с точки зрения внедрения в производство, и поэтому в настоящее время при разработке систем широко используется принцип агрегатизации структуры. *Агрегатизация структуры — это переход от проектирования жесткой структуры к набору автономных функциональных блоков различного назначения, из которых можно компоновать системы произвольной структуры.* Каждый такой блок может самостоятельно выполнять свою специальную функцию, но сконструирован таким образом, что его легко состыковать с другими функциональными блоками агрегатной системы.

Рассмотрим несколько подробнее содержание принципа агрегатизации структур.

При проектировании больших систем автоматизированного управления алгоритмы их функционирования почти неизбежно изменяются и на этапе их разработки, и в процессе внедрения системы, и даже после того, как система уже сдана в промышленную эксплуатацию. Это связано с тем, что: возникает необходимость в улучшении качества управления по мере накопления информации; изменяются условия эксплуатации системы по мере ее «утверждения» на объекте; одни функции могут получить дальнейшее развитие, а другие, наоборот, потерять свою значимость; возможно изменение самого объекта управления, так как в ряде случаев объект управления и система управления проектируются одновременно.

Отсюда и возникает необходимость трансформировать структуру системы, наращивать, если это надо, отдельные устройства, увеличивать их информативную или вычислительную мощность, иметь широкую номенклатуру функциональных блоков, в пределах которой можно строить и развивать структуру систем. Все эти возможности, разумеется, в определенных пределах предоставляет использование агрегатированных комплексов технических средств.

Все функциональные блоки, входящие в агрегатную систему средств, являются совместимыми: по физическим параметрам информационных, управляющих сигналов; конструкциям используемых элементов и узлов агрегатной системы; организации порядка обмена информацией между блоками (по алгоритму связи) в различных режимах связи.

По функциональному назначению все агрегатные устройства обычно делятся: на 1) устройства центрального управления и обработки информации (процессоры); 2) устройства, обеспечивающие информационные и управляющие

связи системы (сбор и вывод информации) с объектом; 3) устройства, обеспечивающие информационные и управляющие связи системы с обслуживающим персоналом; 4) устройства хранения (запоминания) информации; они разделяются на внутренние (оперативная и постоянная память) и внешние (память большой емкости на магнитных барабанах, лентах, дисках); 5) устройства выхода на внешние (внесистемные) линии связи; 6) устройства, обеспечивающие обмен информацией между функциональными блоками системы; 7) устройства ввода информации с внешних носителей (перфоленты, перфокарты, печатающие устройства) и вывода на них.

Анализ основных агрегатированных комплексов технических средств, используемых при построении структуры современных больших автоматизированных систем управления, проведен в работе [50]. К ним относятся: агрегатная система средств вычислительной техники (АСВТ); единая система электронных вычислительных машин (ЕС ЭВМ); агрегатная система средств телемеханики (АССТ); комплекс технических средств для локальных информационно-управляющих систем (КТС ЛИУС). Одновременно с указанными агрегатными системами, получившими уже широкое внедрение, развиваются новые: система малых электронно-вычислительных машин (СМ ЭВМ); система микропроцессоров (СМП); АСКР — агрегатная система средств контроля и регулирования. Рассмотрим кратко характерные особенности основных агрегатных систем.

Агрегатная система средств вычислительной техники. Данная система представляет собой набор агрегатных устройств, предназначенных для создания информационных и управляющих систем для различных отраслей промышленности. В частности, на основе этого комплекса могут быть созданы автоматизированные системы управления для технологических процессов, цехов, электростанций, транспорта, а также автоматизированные системы массового обслуживания и др.

АСВТ является открытой системой в том смысле, что она допускает последовательную модернизацию и наращивание, замену устаревшей аппаратуры на более совершенную, изменение и усовершенствование внутреннего математического обеспечения, позволяет компоновать системы с различными характеристиками и определенным сочетанием технических параметров по объему и виду принимаемой информации, способам и производительности ее обра-

ботки, а также способам формирования выходной информации.

Вся номенклатура устройств АСВТ по своему функциональному назначению делится на классы: а) устройства центрального управления и переработки информации (универсальные и специализированные процессоры); б) устройства хранения информации (оперативные, постоянные, полупостоянные ЗУ, накопители на магнитных лентах, дисках, барабанах); в) устройства связи с объектом, обеспечивающие связь информационной части системы с датчиками и исполнительными механизмами; г) устройства связи с оператором (пульта оператора, различные регистраторы, устройства сигнализации и индикации, мнемосхемы, печатающие механизмы, вводно-выводные устройства на перфокартах, перфолентах, считыватели графической информации, графопостроители и т. д.); д) устройства выхода на внешние линии связи, обеспечивающие подключение системы к телеграфным, радиорелейным, телефонным и другим линиям связи; е) устройства внутрисистемной связи, обеспечивающие согласование различных устройств, входящих в систему.

Информационная совместимость в АСВТ реализуется унификацией единицы передаваемой информации между блоками. В качестве такой единицы принят байт, равный восьми двоичным разрядам; два байта образуют полу-слово, из двух полуслов формируется слово.

Физическая совместимость реализуется регламентацией параметров сигналов и унификацией внутрисистемных связей. Таких унифицированных связей в АСВТ имеется три типа. Каждый тип связи называется рангом. Ранги в зависимости от числа и назначения линий связи делятся на сопряжения. Связи ранга 1 предназначены для осуществления обмена информацией между процессором и оперативной памятью, универсальным процессором и специализированным и отличаются более интенсивным обменом. Передача информации производится параллельно. Максимальная длина связей 1—10 м. Моменты выдачи информации в магистраль обмена определяются специальной управляющей схемой (рис. 8-1), входящей в состав связи ранга 1 (УС-1).

Связи ранга 2 предназначены для соединения устройств с менее интенсивным характером обмена информацией. С помощью этих связей осуществляется, например, соединение процессора с устройствами ввода-вывода. Каждый

байт снабжается дополнительными служебными сигналами, определяющими его назначение. Связи ранга 2 подразделяются на сопряжения 2А, 2В, 2К. Сопряжение 2А имеет,

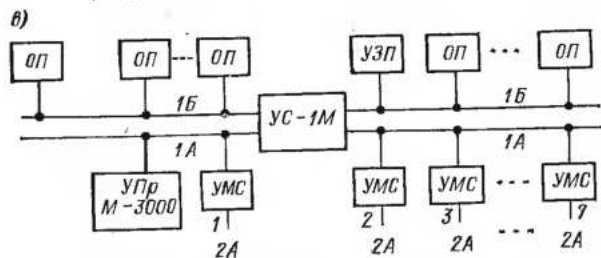
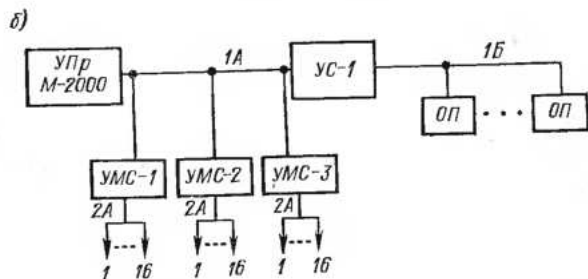
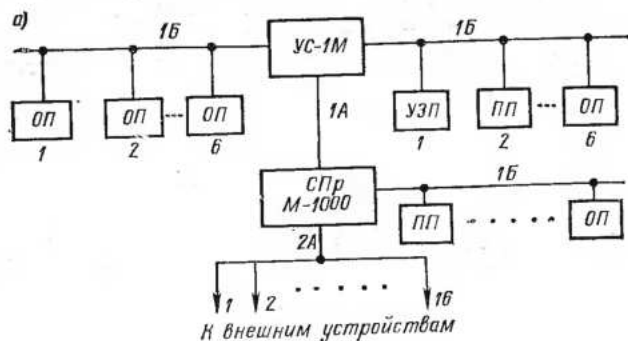


Рис. 8-1. Структура комплексов АСВТ-Д: а — М-1000; б — М-2000; в — М-3000

как и связи ранга 1, магистральную структуру. Сопряжение 2В предназначено для организации радиального соединения устройств. В каждый момент к связи может быть подключено одно из внешних устройств. Сопряжение 2К используется в машинах серии АСВТ-М (М-6000). Обмен информацией по связям ранга 2 производится по принципу

запрос—ответ с помощью параллельно-последовательных кодов.

Связи ранга 3 предназначены для передачи информации на большие расстояния (до 15 км). Способ передачи — последовательный. Единицей передаваемой информации является двоичный разряд (бит). Канал передачи — трехпроводная специальная кабельная линия. В качестве устройств централизованного управления и переработки информации в АСВТ-Д (на дискретных элементах) используются два типа процессоров: специализированные и универсальные.

Специализированный процессор — процессор, обеспечивающий обработку сравнительно небольших массивов информации. Характерная его особенность — совмещение функций обработки информации с функциями обмена информацией между устройствами ввода-вывода.

Универсальный процессор — процессор, предназначенный для решения сложных задач управления (оптимизация, технико-экономическое и оперативно-производственное планирование).

Связь универсального процессора с оперативными запоминающими устройствами осуществляется посредством сопряжений 1А и 1Б, а с устройствами ввода-вывода информации — через мультиплексные каналы, выходящие на сопряжение 2А. Каждый мультиплексный канал обеспечивает подключение до 16 устройств ввода-вывода. В состав АСВТ-Д входят специализированный процессор (М-1000) и универсальные процессоры М-2000 и М-3000.

Возможные схемы соединения специализированных и универсальных процессоров с устройствами оперативной памяти и с устройствами ввода-вывода показаны на рис. 8-1, где 1А, 1Б, 2А — сопряжения рангов 1 и 2; УС-1М — устройство связи ранга 1; СПр — специализированный процессор; ОП, ПП — оперативная и постоянная память; УМС — устройство мультиплексной связи; УС-1 — устройство связи ранга 1; УПр — универсальный процессор; УЗП — устройство защиты памяти.

В АСВТ-Д предусмотрена программная совместимость универсальных процессоров. Специализированный процессор М-1000 программно несовместим с другими моделями процессоров.

Система АСВТ-М (на интегральных микросхемах) включает в себя: а) УВМ (комплексы) М-4000, М-5000, М-6000, М-7000; б) управляющие микропрограммные автоматы

М-6010, М-40. Большой набор команд, современная система математического обеспечения, наличие средств обработки алфавитно-цифровой информации позволяет формировать на базе машин системы АСВТ-М сложные автоматизированные системы управления производством, а также использовать их для обработки данных в вычислительных центрах предприятия.

Единая система электронных вычислительных машин. Данная система машин представляет собой ряд машин, построенных по единому принципу работы, имеющих общую логическую и техническую структуру в том смысле, что применяется единая номенклатура внешних устройств и унифицированная элементная и конструктивно-технологическая база. Ряд машин, входящих в ЕС ЭВМ, включает в себя вычислительные машины различной производительности (от 10 до 1500 тыс. опер/с (с объемом оперативной памяти от 8 до 2048 тыс. байт. В ЕС ЭВМ предусматривается шесть моделей: ЕС-1010, ЕС-1020, ЕС-1030, ЕС-1040, ЕС-1050, ЕС-1060.

Все модели системы отличаются рядом факторов, выдвигающих их в число наиболее современных средств вычислительной техники: а) развитой логической структурой системы; б) универсальным набором операций, обеспечивающим эффективный сбор и обработку информации самого различного характера; в) наличием программных средств контроля за ходом вычислительного процесса, а также весьма совершенных средств управления, позволяющих организовать различные режимы работы; г) наличием разнообразной номенклатуры внешних устройств машины и стандартных процедур для управления этими устройствами, а также для организации их связей с ЭВМ; д) модульным принципом построения как технических, так и программных средств системы; е) возможностью создания систем с переменной структурой и объединения машин в вычислительные комплексы; ж) развитым математическим обеспечением с универсальным набором средств программирования.

На базе ЕС ЭВМ могут быть созданы системы, хорошо приспособленные для автоматизации решения разнообразных научно-технических задач, а также для реализации информационных задач, связанных со сбором большого числа данных, вводимых персоналом, и сложными способами их обработки. Такие машины могут быть использованы с наибольшей эффективностью не для непосредствен-

ного управления технологическим процессом, а в качестве управляющего звена более высокого ранга в иерархической структуре управления сложными технологическими процессами, имеющего в своем подчинении машины нижнего яруса, например, типа М-6000. Это связано с тем, что в составе ЕС ЭВМ отсутствуют технические средства непосредственной связи с объектом управления, а в составе АСВТ такие средства представлены достаточно широко и разнообразно.

Агрегатная система средств телемеханики. Комплекс АССТ предназначен: для телемеханической передачи управляющей, измерительной и сигнальной информации в системах диспетчерского управления и контроля, а также в АСУ, разрабатываемых для различных отраслей народного хозяйства; для управления технологическими процессами на базе вычислительной техники с повышенными требованиями по надежности, когда такая техника используется в нижних ярусах иерархии с прямым управлением исполнительными механизмами технологического оборудования.

АССТ представляет собой набор функциональных блоков (преобразователи кодов, блоки повышения достоверности, коммутаторы, телеизмерительные преобразователи, блоки сопряжения с печатающей машинкой, перфоратором, фотосчитывателем, цифровым индикатором, блоки памяти, устройства синхронизации и др.), объединяемых унифицированными связями и рассчитанных на разработку: а) телемеханических устройств приема и передачи оперативно-технологических и производственно-статистических данных; б) устройств, предназначенных для связи с обслуживающим персоналом; в) устройств сопряжения функциональных блоков с технологическими объектами; г) простейших устройств обработки информации, осуществляющих запоминание, логические и арифметические преобразования и вывод; д) устройств сопряжения средств телемеханики со средствами вычислительной техники (АСВТ).

Комплекс технических средств для локальных информационно-управляющих систем. При автоматизации технологических процессов возникает необходимость в создании местных управляющих средств в силу нецелесообразности загрузки центрального процессора выполнением тех или иных управляющих функций ограниченного характера. В ряде случаев централизация всех функций управления вообще оказывается неэффективной. Это и определяет необходимость в разработке комплекса технических средств

КТС ЛИУС для создания систем либо с децентрализованной структурой, либо построенных по иерархическому принципу с центральным управляющим звеном в верхнем ярусе управления и местными средствами управления в нижнем.

Этот комплекс включает в себя широкую номенклатуру средств, обеспечивающих: получение информации от датчиков различной природы, устанавливаемых на объекте; преобразование сигналов датчика по форме; ввод и коммутацию входной информации для устройств обработки и хранения информации; хранение и обработку информации; вывод информации и преобразование ее в необходимую форму; использование выходных данных при помощи индикаторов, различных регистраторов, исполнительных органов и т. д.

Комплекс КТС ЛИУС включает в себя также средства, обеспечивающие его совместимость с комплексами АСВТ, АССТ, что дает возможность применять средства АСВТ для координации нескольких информационно-управляющих систем на базе КТС ЛИУС, т. е. строить иерархические структуры, о которых говорилось выше.

Рассмотренные агрегатные системы позволяют: а) создавать различные структуры многопроцессорных и многомашинных вычислительных систем с целью повышения, например, их надежности или производительности; б) создавать структуры систем сбора и передачи информации для различных систем объектов управления, как сосредоточенных, так и рассредоточенных в пространстве; при этом используются унифицированные структуры связей, рекомендуемые в выбранной системе агрегатных средств; в) расширять структуры по мере усложнения функций управления и контроля (наращивать объемы памяти, варьировать число и состав устройств ввода-вывода, изменять объем входной и выходной информации, обрабатываемой и формируемой системой управления, и др.).

Агрегатный подход при проектировании БСУ позволяет для решения поставленной перед системой задачи построить ряд структур, различных по своим свойствам, по алгоритму управления, по топологии их внутренних связей. Выбор наилучшей из структур в соответствии с выбранными показателями эффективности является задачей структурного синтеза.

8.2. ПОСТАНОВКА ЗАДАЧИ СТРУКТУРНОГО СИНТЕЗА

При определении структуры БСУ производятся: 1) выбор задач управления, возлагаемых на технические средства; 2) выбор алгоритмов их реализации; 3) формирование общей структуры системы и распределение выбранных задач по узлам и уровням системы; 4) определение комплекса технических средств в узлах системы и их взаимосвязей.

Перечисленные выше этапы создания БСУ взаимно связаны, и задачи каждого из них решаются с учетом ресурсов, выделяемых на создание системы. Выбранная структура системы считается оптимальной, если достигается максимум (минимум) выбранного показателя эффективности, отражающего основные свойства системы с точки зрения выполнения поставленных задач.

На этапе структурного синтеза функции системы представляются в виде совокупности взаимосвязанных задач, которые, в свою очередь, могут быть разбиты на совокупности операций. По известным характеристикам операций и их взаимосвязей могут быть получены соответствующие характеристики задач.

При формализации взаимосвязей между функциями обычно учитываются порядок следования операций и их длительности (временные связи), а также объемы обмениваемой информации (объемные связи). Формализация взаимосвязей обычно производится на основе теории графов. Существующие постановки задачи синтеза структуры БСУ могут различаться по следующим условиям: 1) задано или подлежит выбору множество задач; 2) заданы или подлежат выбору взаимосвязи между задачами (временные, объемные, объемно-временные); 3) заданы или подлежат выбору элементы системы (комплекс технических средств); 4) учитывается или не учитывается территориальное расположение элементов; 5) заданы или подлежат выбору связи между элементами системы; 6) учитывается или не учитывается возможность выполнения задачи в нескольких элементах (распараллеливание выполнения задачи).

Одновременно с этим уже формализованные постановки задачи структурного синтеза могут различаться: а) видом показателей эффективности; б) типом учитываемых характеристик элементов; в) видом ограничений, накладываемых на учитываемые ресурсы (временные, технико-экономические и т. д.). При решении задачи структурного синтеза находят применение различные модели и методы. Широкое

распространение получили модели математического, и особенно дискретного, программирования в силу комбинаторного характера многих задач. Рассмотрим типовую постановку задачи синтеза структуры системы управления сложным объектом [52].

Считаем, что предполагаемая топология размещения возможных узлов (центров) управления известна, функции управления перечислены в виде последовательности задач, которые необходимо распределить между узлами управления. При построении структуры системы управления необходимо распределить функции (задачи) ($i = 1, 2, \dots, I$) между узлами управления ($j = 1, 2, \dots, Y$), выбрать алгоритмические способы реализации функций ($k = 1, 2, \dots, k_i$), типы технических средств в узлах системы ($l = 1, 2, \dots, L$) и вариант организации канала связи между узлами. При этом должны учитываться такие технико-экономические характеристики системы, как затраты на создание A и эксплуатацию B , оперативность, которая характеризует длительность выполнения цикла управления T , надежность системы P , масса W , энергопотребление E и др.

Для формализации постановки задачи введем переменные: $x_{ijkl} = 1$, если i -я задача решается по k -му варианту в j -м узле при помощи l -го технического средства; $x_{jl} = 1$, если j -й узел оборудуется l -м техническим средством; $x_{jj'} = 1$, если необходимо создать канал связи между узлами j и j' ; $x_{ikjl} = x_{jl} = x_{jj'} = 0$ — в противном случае. Тот факт, что каждый вариант построения структуры должен включать в себя лишь один способ распределения задач по узлам системы и один способ выполнения каждой задачи, учитывает ограничение

$$\sum_{k, l, i} x_{ijkl} = 1, \quad i = 1, 2, \dots, I.$$

Переменные x_{jl} , $x_{jj'}$ зависят от x_{ikjl} и используются для удобства записи аналитических выражений расчета различных характеристик вариантов структуры:

$$x_{jl} = \begin{cases} 1, & \text{если } \sum_{i, k} x_{ikjl} \geq 1; \\ 0, & \text{если } \sum_{i, k} x_{ikjl} = 0; \end{cases}$$

$$x_{jj'} = \begin{cases} 1, & \text{если } \sum_{i, k, l, l'} x_{ikjl} x_{ik'j'l'} \geq 1; \\ 0, & \text{если } \sum_{i, k, l, l'} x_{ikjl} x_{ik'j'l'} = 0. \end{cases}$$

Для определения характеристик вариантов структуры введем следующие обозначения: A_l — стоимость технического средства l либо затраты на разработку и изготовление перспективных средств; $A_{jj'}$ — стоимость создания канала связи, включающая в себя стоимость технических средств приема и передачи информации между узлами j и j' ; A_{ikjl} — затраты на разработку информационного и математического обеспечения решения i -й задачи в k -м варианте в j -м узле при наличии l -го технического средства; B_{ikjl} — эксплуатационные затраты на решение i -й задачи в k -м варианте в j -м узле при наличии l -го технического средства; $B_{ij(i+1)j'}$ — затраты на передачу информации от i -й задачи, решаемой в j -м узле, к $(i+1)$ -й задаче, решаемой в j' -м узле; t_{ikjl} — время решения i -й задачи k -м способом в j -м узле при наличии l -го технического средства; $t_{ij(i+1)j'}$ — время передачи информации от i -й задачи, решаемой в j -м узле, к $(i+1)$ -й задаче, решаемой в j' -м узле; P_{ikjl} — надежность решения задачи (технического средства), определяемая либо вероятностью безотказной работы, либо функцией ее; $P_{ij(i+1)j'}$ — надежность передачи информации (канала связи), определяемая как P_{ij} и P_{ikj} ; W_l — масса технического средства; E_l — энергопотребление технического средства.

Используя введенные обозначения, выпишем формулы для определения характеристик вариантов структуры.

Капитальные затраты A включают в себя стоимость технических средств в узлах, стоимость создания каналов связи между узлами системы, затраты на разработку алгоритмов выполнения задач:

$$A = \sum_{i, l} A_l x_{ijl} + \sum_{j, j'} A_{jj'} x_{jj'} + \sum_{i, k, j, l} A_{ikjl} x_{ikjl}.$$

Эксплуатационные затраты B включают в себя затраты на выполнение задач управления и затраты на передачу информации между узлами системы:

$$B = \sum_{i, k, j, l} B_{ikjl} x_{ikjl} + \sum_{k', l, j, j', l'} B_{ij(i+1)j'} x_{ikjl} x_{(i+1)k'j'l'}.$$

Время выполнения цикла управления, характеризующее оперативность управления, вычисляется аналогично:

$$T = \sum_{i, k, j, l} t_{ikjl} x_{ikjl} + \sum_{k', l, j, j', l'} t_{ij(i+1)j'} x_{ikjl} x_{(i+1)k'j'l'}.$$

При определении надежности системы будем считать систему невосстанавливаемой, в результате чего ее надежность, характеризуемая вероятностью безотказной работы элементов, равна произведению вероятностей безотказной работы элементов. Поэтому

$$\bar{P} = \sum_{i, j, k, l} \bar{P}_{ijkl} x_{ijkl} + \sum_{k', i, j, k, l, i', j', l'} \bar{P}_{ij(i+1)j'} x_{ikjl} x_{(l+1)k'j'l'},$$

где $\bar{P}$, $\bar{P}_{ijkl}$, $\bar{P}_{ij(i+1)j'}$ — логарифмы соответствующих величин.

Масса и энергопотребление технических средств в узлах системы

$$W_j = \sum_{i, k, l} W_{ijkl} x_{ijkl};$$

$$E_j = \sum_{i, k, l} E_{ijkl} x_{ijkl}, \quad j = 1, 2, \dots, J.$$

При решении задачи синтеза структуры любая из указанных характеристик может быть выбрана в качестве показателя эффективности (в зависимости от цели и назначения системы), который оптимизируется, а другие учитываются в ограничениях. Поэтому в общем виде задача оптимизации структуры автоматизированной системы управления выглядит следующим образом:

$$\begin{aligned} & \min (\max) f(x_{ijkl}, x_{jl}, x_{jj'}); \\ & f_s(x_{ijkl}, x_{jl}, x_{jj'}) \leq D_s, \quad s = 1, 2, \dots, S; \\ & \sum_{k, j, l} x_{ijkl} = 1, \quad i = 1, 2, \dots, I; \\ & x_{ijkl} = (0; 1); \quad x_{jl} = (0; 1); \quad x_{jj'} = (0; 1). \end{aligned}$$

Из приведенной постановки задачи следует, что она является моделью дискретного программирования и носит комбинаторный характер. При решении таких задач возникают трудности принципиального характера, а именно: необходимо исключать явный перебор всех допустимых решений и стремиться к эффективному частичному перебору сравнительно малого числа допустимых вариантов решения соответствующей задачи и неявному перебору остальных. Рассмотрим далее основные подходы к построению вычислительных алгоритмов решения подобных задач.

8.3. ОСНОВНЫЕ ПРИНЦИПЫ ПОСТРОЕНИЯ АЛГОРИТМОВ СТРУКТУРНОЙ ОПТИМИЗАЦИИ

При решении задач структурной оптимизации, сформулированных в виде задач дискретного программирования, используются три группы методов: 1) методы отсечения; 2) комбинаторные методы; 3) приближенные методы.

Методы отсечения. В случае, если целевая функция и функции ограничений являются линейными, задача является частным случаем задач линейного целочисленного программирования. Поэтому центральной идеей этой группы методов является использование в их вычислительной схеме алгоритма решения соответствующей задачи линейного программирования. При этом учет целочисленности приводит к появлению новых дополнительных ограничений и расширению общей системы ограничений.

В 1958 г. впервые Р. Гомори удалось построить универсальный и реализуемый в вычислительном отношении алгоритм, в котором последовательно была развита идея методов отсечения. Позднее ряд других авторов разработали различные модификации методов Р. Гомори, учитывающие специфику различных ограничений задачи. Следует отметить, что в общем случае эффективность решений задач данного класса методами отсечений сравнительно низкая, так как с увеличением числа переменных и ограничений задачи замечено увеличение числа итераций, необходимых для нахождения оптимального решения исходной задачи.

Алгоритмы, строящиеся на использовании идеи методов отсечения, не гарантируют получения допустимого целочисленного решения до самой последней итерации, поэтому их эффективное использование на практике зачастую затруднительно, особенно для задач большой размерности.

Комбинаторные методы. Методы решения задач целочисленного линейного программирования в классе комбинаторных методов базируются на максимальном учете характера задач и конечности множества вариантов их решения. В их вычислительных схемах используется идея частичного перебора вариантов решения задачи. Это достигается путем отбрасывания некоторых подмножеств вариантов, которые согласно известным свойствам оптимального решения заведомо могут считаться неперспективными.

Отличительной особенностью комбинаторных методов является и то, что во многих из них вообще не используется

решение задачи линейного программирования, соответствующей исходной целочисленной задаче линейного программирования.

Таким образом, для большинства комбинаторных методов не требуется специальных доказательств их конечности, что и способствовало широкому размаху применений методов этого типа на практике. Большая группа комбинаторных методов базируется в идейном отношении на достаточно общей схеме методов, которые объединены под одним названием — методы «ветвей и границ». В настоящее время схема метода «ветвей и границ» широко используется на практике и методы, строящиеся на ее использовании, заняли, пожалуй, центральное место среди комбинаторных методов.

Приближенные методы. Одной из основных причин бурного развития этой группы методов следует назвать сравнительную сложность реализации точных методов, а зачастую и просто невозможность применения последних для задач большой размерности и для задач, время решения которых ограничено. Другим стимулом, способствующим на практике развитию приближенных методов, является то, что для многих практических задач в значительной мере точные решения не обеспечиваются из-за малой достоверности исходных данных.

Среди приближенных методов решения целочисленных задач линейного программирования следует выделить применение различного рода эвристических алгоритмов, основанных: а) на методах, построенных на использовании случайного поиска; б) методах, сочетающих случайный поиск с идеей локальной оптимизации; в) методах, вычислительные схемы которых строятся на максимальном учете специфики конкретных типов задач, и др.

Учитывая широкое практическое применение методов «ветвей и границ» для решения комбинаторных задач, в класс которых в основном попадают задачи структурной оптимизации, рассмотрим алгоритмическую схему метода «ветвей и границ».

В настоящее время имеется уже много разработанных модификаций этого метода, которые могут быть успешно использованы для решения различных типов задач дискретного программирования. В дальнейшем для определенности рассмотрим алгоритм метода «ветвей и границ» будем считать, что в поставленной задаче оптимизации ищется минимум целевой функции. В основу метода «ветвей и границ» положены следующие этапы.

1. Определение нижней оценки целевой функции на множестве допустимых решений X_0 . Под нижней границей оценки целевой функции $f(x)$ понимается такое значение $V(X_0)$, при котором для всех допустимых решений $x = (x_1, x_2, \dots, x_n)$ имеет место

$$f(x^{\min}) \geq V(X_0), x \in X_0. \quad (8-1)$$

Оценка $V(X_0)$ должна также удовлетворять следующему свойству: оценка по множеству допустимых решений должна быть не хуже оценки по любому входящему в него подмножеству, т. е.

$$V(X_0) \leq V(X_1), X_1 \subset X_0. \quad (8-2)$$

2. Ветвление, т. е. построение дерева решений или разбиение всего множества решений на подмножества (рис. 8-2) в соответствии с выбранным признаком, причем $X_0 = \bigcup X_{i_1}, X_{i_1} = \bigcup X_{i_1 i_2} \dots$. На первом шаге при определенном выше правиле определения оценок ищется

$$\min_{i_1} V(X_{i_1}) = V(X_{r_1}).$$

На втором шаге производится ветвление подмножества X_{r_1} и вычисление соответствующих оценок $V(X_{r_1, i_2})$ на подмножествах X_{r_1, i_2} . Далее ищется

$$\min_{i_2} V(X_{r_1, i_2}) = V(X_{r_1, r_2}).$$

Процесс ветвления продолжается до тех пор, пока полностью не будет определено решение $x^{(0)}$, «подозреваемое» на оптимальность; в этом случае выделенное подмножество решений включает в себя всего только один вариант $x^{(0)}$.

3. Определение оптимальности решения $x^{(0)}$. Для тех случаев, когда оценка $V(X_0)$ построена таким образом, что позволяет определить точную нижнюю границу для значений целевой функции $f(x)$ на множестве

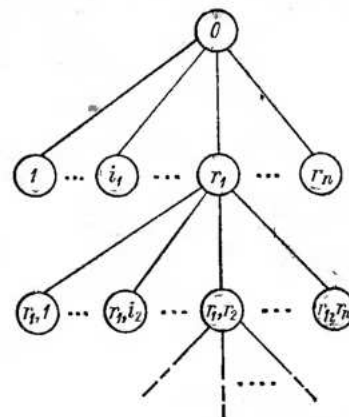


Рис. 8-2. Построение дерева решений в методе «ветвей и границ»

допустимых решений, то найденное решение $x^{(0)}$ является строго оптимальным решением $x^{\min}$.

Если построенная оценка $V(X_0)$ является только грубой оценкой нижней границы для значений целевой функции $f(x)$, т. е. неравенство (8-1) является строгим, то решение $x^{(0)}$ в общем случае является приближенно оптимальным решением (субоптимальным), хотя может быть и строго оптимальным.

В задачах дискретного программирования, носящих комбинаторный характер, чаще используются грубые оценки, ибо они проще строятся и вычисляются, а поэтому получение в случае применения грубой оценки решения $x^{(0)}$ еще не гарантирует его оптимальности.

В случае применения точных оценок $V(X_0)$ получение оптимального решения $x^{(0)}$ заканчивает работу алгоритма метода «ветвей и границ». При использовании грубых оценок возможно возникновение такой ситуации, когда полученное значение $V(x^{(0)})$ оказывается хуже, чем значения оценок на некоторых неразветвленных подмножествах построенного дерева решений. В этом случае обычно организуется дополнительное ветвление, позволяющее улучшить значение $x^{(0)}$, алгоритм которого сводится к следующему.

Пусть уровень последнего ветвления, на котором определено решение $x^{(0)} = (x_{r_1}, x_{r_2}, \dots, x_{r_n})$, есть уровень с номером $(n-1)$. Обозначим значение $V(x^{(0)})$ через R и назовем рекордом. Поднимаемся на уровень с номером $(n-2)$ и ищем

$$\min_{i_{n-2}} V(X_{r_1, r_2, \dots, r_{n-3}, i_{n-2}}), \quad i_{n-2} \neq r_{n-2}.$$

Если оказывается, что

$$R \leq \min_{i_{n-2}} V(X_{r_1, r_2, \dots, r_{n-2}}),$$

то это означает, что на данном уровне нет подмножеств, которые могли бы дать при дополнительном ветвлении решение лучшее, чем найденное $x^{(0)}$, в силу ранее принятого свойства оценки (8-2). Поэтому переходим на уровень $(n-3)$. Если получили, что $R > \min_{i_{n-2}} V(X_{r_1, r_2, \dots, r_{n-2}})$,

$$i_{n-2} \neq r_{n-2},$$

то определяем s_{n-2} , на котором достигается

$$V(X_{r_1, r_2, \dots, r_{n-3}, s_{n-2}}) = \min_{i_{n-2}} V(X_{r_1, r_2, \dots, r_{n-3}, i_{n-2}}),$$

и производим дополнительное ветвление подмножества $X_{r_1, r_2, \dots, r_{n-3}, s_{n-2}}$ до тех пор, пока не происходит ухудшения получаемых оценок по сравнению с найденным значением рекорда R . В противном случае переходим на уровень $(n-3)$. Если в процессе дополнительного ветвления определен новый рекорд, который оказывается лучше, чем R ($R_1 < R$), то рассмотренная выше процедура повторяется для нового рекорда R_1 . Если $R_1 \geq R$, то переходим на уровень с номером $(n-3)$, сохраняя старое значение рекорда R , и рассмотренная выше процедура улучшения значения $x^{(0)}$ повторяется начиная с уровня $(n-3)$ и далее. Данный алгоритм улучшения значения $x^{(0)}$ при использовании грубых оценок заканчивает работу, если найденное значение рекорда оказывается не хуже (меньше либо равно), чем значения оценок на всех неразветвленных подмножествах построенного дерева решений.

В общем случае алгоритм дополнительного ветвления, реализуемый в возвратно-поступательном режиме, может привести к полному перебору вариантов и чем грубее оценка, тем больше вероятность приблизиться к полному перебору при поиске оптимального решения. Однако можно ограничиться поиском приближенного оптимального решения, если ввести оценку точности приближенного оптимального решения ϵ . В этом случае ветвление любого подмножества X не проводится, если

$$|R - V(X)| \leq \epsilon.$$

Рассмотренная схема применения метода «ветвей и границ» является достаточно общей. Однако при реализации решения конкретных задач дискретного программирования требуется, исходя из максимального учета специфики решаемой задачи, конкретизировать способы вычисления оценок и методы ветвления. Удачное решение этих вопросов во многом способствует повышению эффективности вычислительного процесса, основанного на использовании данного метода. Для иллюстрации данного метода рассмотрим примеры.

Пример 8-1. Распределение задач по узлам управления. Задано множество задач, реализуемых в АСУ ($i = 1, 2, \dots, I$); задано множество узлов системы управления ($j = 1, 2, \dots, J$). Необходимо так распределить задачи по узлам системы, чтобы достигнуть максимального значения эффективности, не выходя из области допустимых ограничений. Пусть c_{ij} — затраты на реализацию i -й задачи в j -м узле, t_{ij} — время решения задачи в j -м узле.

Введем дополнительную переменную x_{ij} ; $x_{ij} = 1$, если i -я задача выполняется в j -м узле, и $x_{ij} = 0$ — в противном случае. Оптимизация распределения задач по узлам может производиться по одной из следующих целевых функций:

$$\min_{x_{ij}} \sum_{i=1}^I \sum_{j=1}^J c_{ij} x_{ij}; \quad (8-3)$$

$$\min_{x_{ij}} \sum_{i=1}^I \sum_{j=1}^J t_{ij} x_{ij}; \quad (8-4)$$

$$\min_{x_{ij}} \left[\max_i \left(\sum_{j=1}^J t_{ij} x_{ij} \right) \right]. \quad (8-5)$$

Функция (8-3) соответствует минимизации затрат, функция (8-4) — минимизации общего времени решения задачи, а функция (8-5) — минимизации максимального времени решения задачи в узле. При этом могут учитываться следующие ограничения: а) связи между задачами (обычно задаются графом $G(I)$, где I — множество задач); б) связи между узлами, в которых находятся элементы системы (обычно задаются графом $G(J)$, где J — множество узлов);

в) ограничение на общее время решения всех задач

$$\sum_{i=1}^I \sum_{j=1}^J t_{ij} x_{ij} \leq T_3,$$

если минимизируются затраты, либо ограничение на общие затраты по реализации задач в АСУ

$$\sum_{i=1}^I \sum_{j=1}^J c_{ij} x_{ij} \leq c_3,$$

если минимизируется время решения всех задач;

г) ограничение на загрузку каждого узла

$$\sum_{i=1}^I \lambda_i t_{ij} x_{ij} \leq \rho_{3j}, \quad j = 1, 2, \dots, J,$$

где λ_i — интенсивность поступления i -й задачи на решение; ρ_{3j} — допустимая загрузка j -го узла;

д) если каждая задача решается только в одном узле системы, то

$$\sum_{j=1}^J x_{ij} = 1, \quad i = 1, 2, \dots, I.$$

В зависимости от того, как выбрана целевая функция и какие ограничения учитываются, возникает ряд частных постановок задачи оптимального распределения задачи по узлам.

Минимизация общих затрат (общего времени) при ограничениях на загрузку каждого из узлов и при условии, что каждая задача решается

только в одном узле системы, т. е.

$$\min \sum_{i=1}^I \sum_{j=1}^J c_{ij} x_{ij}, \text{ либо}$$

$$\min \sum_{i=1}^I \sum_{j=1}^J t_{ij} x_{ij}, \text{ при}$$

$$\sum_{i=1}^I \lambda_i t_{ij} x_{ij} \leq \rho_{3j}$$

$$\sum_{j=1}^J x_{ij} = 1; \quad j = 1, 2, \dots, J; \quad i = 1, 2, \dots, I.$$

Минимизация общих затрат (общего времени) при ограничениях на общее время (общие затраты) и при условии, что каждая задача решается только в одном узле системы, т. е.

$$\left. \begin{aligned} \min \sum_{i=1}^I \sum_{j=1}^J c_{ij} x_{ij}; \\ \sum_{i=1}^I \sum_{j=1}^J t_{ij} x_{ij} \leq T_3; \\ \sum_{j=1}^J x_{ij} = 1; \quad i = 1, 2, \dots, I; \quad x_{ij} = (0; 1). \end{aligned} \right\} \quad (8-6)$$

В дальнейшем будем рассматривать только задачу (8-6). Дерево ветвления строится следующим образом. Подмножество первого уровня разбиения формируем, фиксируя соответствие первой задачи различным узлам ($X_1, X_2, \dots, X_{j_1}, \dots, X_J$). Множество X_{j_1} включает в себя все варианты, где первая задача решается в узле j_1 , а распределение остальных задач по узлам произвольное. Аналогично, множества второго уровня формируем, фиксируя соответствие второй задачи различным узлам. Множество X_{j_1, j_2} включает в себя все варианты решений, где первая задача решается в узле j_1 , вторая задача — в узле j_2 , а остальные задачи имеют произвольное распределение по узлам и т. д.

Для каждого из подмножеств (вершин дерева) необходимо построить оценки целевой функции и ограничения. Общее выражение оценки целевой функции для множества вариантов $X_{j_1 j_2 \dots j_i \dots j_I}$ в данной задаче может быть построено следующим образом:

$$V_c(X_{j_1 j_2 \dots j_i \dots j_I}) = \sum_{i \leq l} c_{ij_i} + \sum_{i > l} \min_i c_{ij_i}, \quad (8-7)$$

а общее выражение для оценки ограничения может быть построено аналогично:

$$V_t(X_{j_1 j_2 \dots j_i \dots j_I}) = \sum_{i \leq l} t_{ij_i} + \sum_{i > l} \min_i t_{ij_i}, \quad (8-8)$$

где $j_1, j_2, \dots, j_i, \dots, j_I$ — множество узлов системы, закрепленных за соответствующими задачами $1, 2, \dots, i, \dots, I$.

Оценка для функции ограничения (8-8) необходима в данном случае для исключения из процесса ветвления множества заведомо не подходящих вариантов с учетом принятого ограничения на общее время решения задач. Для ускорения процесса поиска используются всякие дополнительные приемы, учитывающие специфику задачи. Рассмотрим их на численном примере. Пусть заданы: матрица стоимостных затрат

$$C = \begin{vmatrix} 3 & 7 & 2 & 2 \\ 4 & 8 & 1 & 3 \\ 5 & 9 & 6 & 2 \\ 6 & 10 & 7 & 1 \\ 7 & 5 & 3 & 1 \end{vmatrix};$$

матрица временных затрат

$$T = \begin{vmatrix} 1,5 & 3 & 2 & 9 \\ 2 & 6 & 5 & 10 \\ 3 & 7 & 6 & 11 \\ 4 & 8 & 7 & 12 \\ 4 & 9 & 8 & 5 \end{vmatrix}.$$

Значение $T_3 = 20$.

Так как каждая задача может решаться только в одном узле системы, то

$$\min_{x_{ij}} \sum_{i=1}^I \sum_{j=1}^J c_{ij} x_{ij} = \sum_{i=1}^I \min_{x_{ij}} \sum_{j=1}^J c_{ij} x_{ij}.$$

Используя это условие и принимая во внимание ограничение по времени, можно исключить из матриц C и T элементы, которые не влияют на выбор оптимального решения. Такими элементами в каждой строке являются элементы, удовлетворяющие одновременно двум следующим условиям:

$$c_{ij} > \min_i c_{ij} = c_{is}; \quad t_{ij} > t_{is}.$$

После проведения такого преобразования матрицы $C^{(0)}$ и $T^{(0)}$ будут выглядеть следующим образом:

$$C^{(0)} = \begin{vmatrix} 3 & - & 2 & - \\ 4 & - & 1 & - \\ 5 & 9 & 6 & 2 \\ 6 & 10 & 7 & 1 \\ 4 & - & - & 1 \end{vmatrix}; \quad T^{(0)} = \begin{vmatrix} 1,5 & - & 2 & - \\ 2 & - & 5 & - \\ 3 & 7 & 6 & 11 \\ 4 & 8 & 7 & 12 \\ 4 & - & - & 5 \end{vmatrix}.$$

Принимая во внимание ограничение по времени $T_3 = 20$, можно из матриц $C^{(0)}$ и $T^{(0)}$ также вычеркнуть элементы, при которых всегда будет нарушено ограничение по времени. Распишем для выделения таких элементов данное ограничение в следующем виде:

$$\sum_{i=1}^{r-1} \min_i t_{ij} + t_{rj} + \sum_{i=r+1}^I \min_i t_{ij} \leq T_3. \quad (8-9)$$

Изменяя индекс строки r ($r = 1, 2, \dots, I \dots, I$) и просматривая все элементы этой строки в соответствии с условием (8-9), можно еще

раз преобразовать матрицы $C^{(0)}$ и $T^{(0)}$, вычеркнув такие элементы, если они имеются. Проведем такое преобразование матриц $C^{(0)}$ и $T^{(0)}$; в результате получим:

$$C^{(1)} = \begin{vmatrix} 8 & - & 2 & - \\ 4 & - & 1 & - \\ 5 & 9 & 6 & - \\ 6 & 10 & 7 & - \\ 7 & - & - & 1 \end{vmatrix}; \quad T^{(1)} = \begin{vmatrix} 1,5 & - & 2 & - \\ 2 & - & 5 & - \\ 3 & 7 & 6 & - \\ 4 & 8 & 7 & - \\ 4 & - & - & 5 \end{vmatrix}.$$

Из матриц $C^{(1)}$ и $T^{(1)}$ видно, что исходные матрицы значительно упростились после проведения дополнительных преобразований. Перейдем теперь непосредственно к решению задачи на основе алгоритма ветвей и границ. Дерево решения с учетом построенных оценок (8-7) и (8-8) представлено на рис. 8-3.

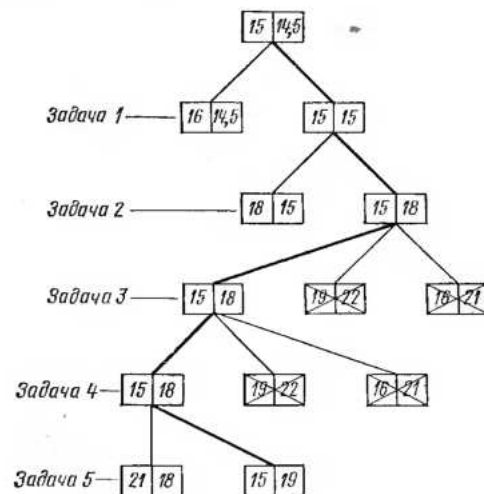


Рис. 8-3. Дерево решений при распределении задач по узлам АСУ

В узлах дерева слева представлена оценка целевой функции, справа — оценка ограничения; зачеркнуты узлы, не удовлетворяющие ограничению. В результате решения получили следующее оптимальное распределение задач по узлам: 1 — 3; 2 — 3; 3 — 1; 4 — 1; 5 — 4 (задача — узел). Жирной чертой на рис. 8-3 представлен путь поиска оптимального решения на дереве вариантов.

Из примера видно, что учет специфики задачи значительно повышает эффективность вычислительной процедуры метода «ветвей и границ». Рассмотренная на примере схема метода «ветвей и границ» легко может быть перенесена на общую постановку задачи, приведенную в § 8-2.

Пример 8-2. Распределение алгоритмов решения по задачам. Данную постановку задачи будем рассматривать на примере выбора

средств повышения достоверности в сложных автоматизированных системах контроля (АСК).

Качество работы любой проектируемой АСК обычно характеризуется показателями, отражающими следующие свойства: достоверность, быстродействие, масса, объем, стоимость. К числу главных

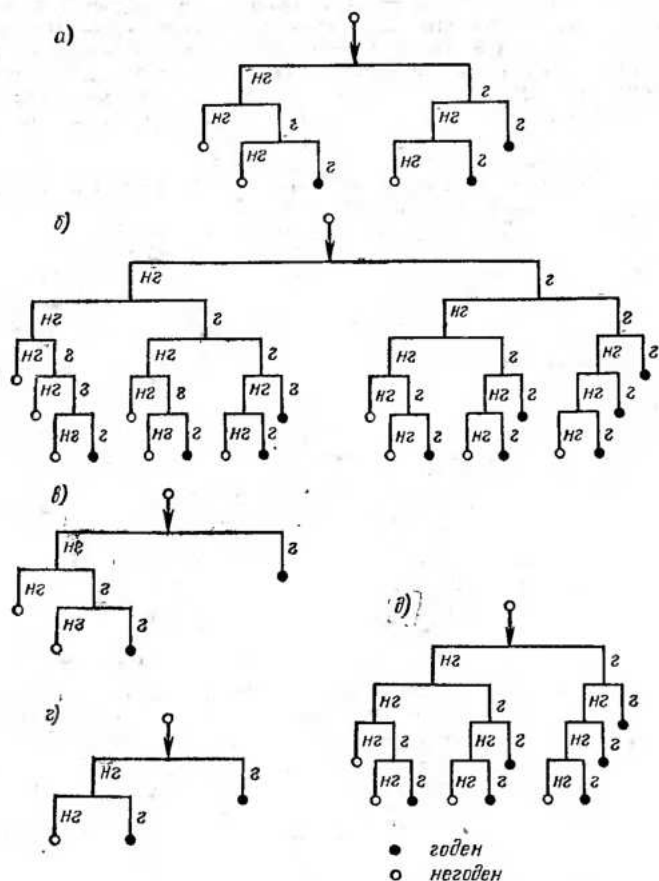


Рис. 8-4. Структуры алгоритмов принятия решений в АСК: а — алгоритм 2 из 3; б — алгоритм 3 из 5; в — усеченный алгоритм 2 из 3 (вариант А); г — усеченный алгоритм 2 из 3 (вариант Б); д — усеченный алгоритм 3 из 5

свойств, характеризующих систему АСК и определяющих ее основное назначение, относят: достоверность и быстродействие контроля.

Среди различных способов повышения достоверности АСК в настоящее время широко используется метод введения программно-логических средств [31], требующий минимальных аппаратных

затрат и позволяющий повышать достоверность результатов контроля при наличии возможных возмущений в работе АСК в виде сбоев и случайных погрешностей в измерительных каналах.

В дальнейшем будем рассматривать системы многопараметрического контроля с двухальтернативной системой принятия решений: параметр в допуске, т. е. «годен», параметр не в допуске, т. е. «негоден». При окончании контроля параметра в зависимости от результата контроля в АСК производится или не производится повторное выполнение программы контроля данного параметра, причем выдача сигнала на очередное повторение программы контроля параметра зависит от алгоритма принятия решения, выбранного в системе программно-логических средств.

Выбор алгоритма принятия решений влияет как на достоверность АСК, так и на ее быстродействие. На рис. 8-4 приведены несколько возможных алгоритмов принятия решений.

В качестве модели влияния случайных возмущений на результаты контроля параметра для сравнительной оценки эффективности введения различных типов алгоритмов принятия решений в АСК была принята информационная модель бинарного канала, которая характеризуется следующими состояниями: $X = (x_1, x_2)$ — множество состояний контролируемого параметра, причем состояние x_1 — параметр «в допуске», x_2 — «не в допуске»; $Y = (y_1, y_2)$ — множество состояний результата контроля параметра, y_1 — параметр «годен», y_2 — параметр «негоден». Данная модель задается вектором априорных вероятностей состояний параметра P_X и матрицей P_{XY} , учитывающей влияние возмущений на работу системы при контроле параметра

$$P_X = \begin{bmatrix} P \\ 1-P \end{bmatrix}; P_{XY} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},$$

где P — вероятность наличия состояния контролируемого параметра «в допуске» в момент его контроля; при этом предполагается, что состояние параметра объекта контроля не изменяется в течение времени его контроля; P_{12} , P_{21} — условные вероятности наличия возмущений в АСК в течение времени одного цикла контроля, приводящие к ошибкам первого и второго рода.

Обозначим: $P_{12} = \alpha$, $P_{21} = \beta$; тогда $P_{11} = 1 - \alpha$, $P_{22} = 1 - \beta$. Из данной информационной модели можно получить выражения для показателей достоверности, в качестве которых выбраны условные апостериорные вероятности $P(X_1/Y_1)$ и $P(X_2/Y_2)$ наличия соответствующего состояния контролируемого параметра при условии, что такое же состояние параметра обнаружено по результатам контроля, определяемые для каждого из рассматриваемых алгоритмов следующим образом:

$$D_F^{(i)} = P^{(i)}(X_1/Y_1) = \frac{P(1 - P_{12}^{(i)})}{P(1 - P_{12}^{(i)}) + (1 - P)P_{21}^{(i)}};$$

$$D_{нг}^{(i)} = P^{(i)}(X_2/Y_2) = \frac{(1 - P)(1 - P_{21}^{(i)})}{(1 - P)(1 - P_{21}^{(i)}) + P P_{12}^{(i)}},$$

где $P_{12}^{(i)}$, $P_{21}^{(i)}$ — эквивалентные условные вероятности ошибок для принятого вида алгоритма ($i = 1, 2, 3, 4, 5$); $D_F^{(i)}$, $D_{нг}^{(i)}$ — показатели достоверности результатов контроля «годен» и «негоден».

Вероятности $P_{12}^{(i)}$, $P_{21}^{(i)}$ могут быть определены в соответствии с i -м алгоритмом принятия решений. Для полного алгоритма 2 из 3 эквивалентные вероятности $P_{12}^{(1)}$ и $P_{21}^{(1)}$ определяются следующим образом: $P_{12}^{(1)} = \alpha^2 + 2\alpha^2(1-\alpha)$; $P_{21}^{(1)} = \beta^2 + 2\beta^2(1-\beta)$.

В силу малости значений α и β ($\alpha \ll 1$, $\beta \ll 1$) приближенные соотношения для $P_{12}^{(1)}$, $P_{21}^{(1)}$ выглядят следующим образом: $P_{12}^{(1)} \approx 3\alpha^2$, $P_{21}^{(1)} \approx 3\beta^2$.

Выбирая в качестве показателя быстроты действия контроля среднее время контроля параметра, отнесенное к времени одной проверки параметра $\bar{\tau}^{(i)}$, для полного алгоритма 2 из 3 получим: $\bar{\tau}^{(1)} \approx 2 + 2\beta + 2\alpha$ ($\alpha - \beta$).

Из анализа соответствующих алгоритмов принятия решений могут быть получены приближенные выражения для эквивалентных вероятностей $P_{12}^{(i)}$, $P_{21}^{(i)}$ и среднего относительного времени контроля параметра, которые приведены в табл. 8-1,

Таблица 8-1

№ п/п	Вид алгоритма	$P_{12}^{(i)}$	$P_{21}^{(i)}$	$\bar{\tau}^{(i)}$
0	Нет алгоритма	α	β	1
1	2 из 3	$3\alpha^2$	$3\beta^2$	$2 + 2\beta + 2P(\alpha - \beta)$
2	3 из 5	$10\alpha^3$	$10\beta^3$	$3 + 3\beta + 3P(\alpha - \beta)$
3	2 из 3 (А)	$2\alpha^2$	β	$1 + (1 - P) + 2P\alpha$
4	2 из 3 (Б)	α^2	2β	$1 - \beta + (1 - P) + P(\alpha + \beta)$
5	3 из 5 (усеченный)	$4\alpha^3$	$6\beta^2$	$2 + 3\beta + (1 - P) + P(2\alpha - 3\beta)$

Проведем сравнительный анализ алгоритмов по основным показателям — достоверности и времени контроля одного параметра. Для этого преобразуем выражения $D_r^{(i)}$ и $D_{nr}^{(i)}$ к следующему виду:

$$D_r^{(i)} = 1 \left/ \left[1 + \frac{1-P}{P} \cdot \frac{P_{21}^{(i)}}{1-P_{12}^{(i)}} \right] \right.;$$

$$D_{nr}^{(i)} = 1 \left/ \left[1 + \frac{P}{1-P} \cdot \frac{P_{12}^{(i)}}{1-P_{21}^{(i)}} \right] \right. \quad (8-10)$$

Из выражений (8-10), учитывая при этом, что для современных АСК $\alpha \leq 10^{-2}$, $\beta \leq 10^{-2}$, $(1-P) < 10^{-2}$, видно, что мин $P_{21}^{(i)}$ соответствует макс $D_r^{(i)}$, а мин $P_{12}^{(i)}$ соответствует макс $D_{nr}^{(i)}$. Итак, сравнение алгоритмов принятия решений по достоверности для одного параметра эквивалентно сравнению по условным вероятностям $P_{12}^{(i)}$, $P_{21}^{(i)}$. Из анализа данных табл. 8-1 можно сделать следующие выводы: 1) для уменьшения среднего времени контроля параметра целесообразно вводить усеченные алгоритмы принятия решений; 2) для повышения достоверности D_r (на параметре) целесообразно использовать полные алгоритмы 2 из 3 и 3 из 5; 3) для повышения достоверности D_{nr} (на параметре) целесообразно использовать усеченные алгоритмы принятия решений. Выбор конкретного типа эффективных алгоритмов (на пара-

метре) зависит от требований к достоверности, к длительности контроля и от значения вероятностных характеристик α , β , P .

В общем случае программа контроля АСК состоит из ряда программ контроля отдельных независимых параметров, обладающих различными характеристиками P и требующих для своей проверки формирования различных трактов обработки и измерения, имеющих различные значения α и β . Так как конечной задачей АСК является определение с требуемой достоверностью фактического состояния контролируемого объекта в течение заданного отрезка времени, то можно сформулировать задачу выбора оптимальной структуры алгоритмов для всего объекта в целом, рассматривая одновременно и достоверность и быстроту действия контроля проектируемой АСК.

Очевидно, что, подбирая тот или иной вид алгоритма для рассматриваемых параметров в зависимости от требуемой достоверности, можно добиться значительной эффективности контроля без усложнения АСК, увеличения ее габаритов, массы, стоимости.

Зная характеристики надежности объекта контроля и АСК, характеристики погрешностей в измерительных каналах, а также требуемые уровни достоверности, можно распределить различные виды алгоритмов по отдельным параметрам таким образом, чтобы обеспечить заданную достоверность контроля за минимальное время контроля.

Прежде чем сформулировать задачу оптимизации, необходимо определить целевую функцию. Было бы естественным принять за целевую функцию один из показателей достоверности, так как последние в первую очередь определяют назначение проектируемых АСК. Однако такая постановка задачи не всегда может быть допустимой, так как полученное в этом случае оптимальное значение достоверности может не соответствовать допустимому уровню достоверности, принятому для данного класса АСК. В связи с этим более целесообразно сформулировать поставленную задачу как обеспечение заданной достоверности при минимальном времени контроля.

Сформулируем более подробно задачу оптимизации. Пусть необходимо синтезировать программу контроля с учетом программно-логических средств, состоящую из N программ контроля отдельных независимых параметров, причем в программе контроля каждого параметра может быть использован один из M возможных вариантов алгоритмов принятия решений.

Пусть i — индекс i -го параметра ($i = 1, 2, \dots, N$); j — индекс j -го способа повышения достоверности (алгоритма), $j = 1, 2, \dots, M$.

Введем переменную x_{ij} , причем $x_{ij} = 1$, если для i -го параметра принят j -й способ; $x_{ij} = 0$ — в противном случае. В качестве показателей достоверности приняты

$$D_r = P(X_1/Z_1) = \prod_{i=1}^N \frac{P_i(1-\alpha_i)}{P_i(1-\alpha_i) + (1-P_i)\beta_i};$$

$$D_{nr} = P(X_2/Z_2) = 1 - \frac{\prod_{i=1}^N P_i - \prod_{i=1}^N P_i(1-\alpha_i)}{1 - \prod_{i=1}^N [P_i(1-\alpha_i) + (1-P_i)\beta_i]},$$

где D_r — апостериорная вероятность того, что значения всех N параметров объекта лежат в поле допуска при условии, что по всем пара-

метрам АСК приняла решение «годен»; $D_{\text{нг}}$ — апостериорная вероятность того, что значение хотя бы одного из параметров лежит вне поля допуска при условии, что хотя бы по одному из них АСК приняла решение «негоден».

Целевая функция для данной задачи оптимизации будет выглядеть

$$\bar{T} = \min \prod_{i=1}^N \prod_{j=1}^M x_{ij} \bar{r}^{(i)}(j),$$

а ограничения:

$$1) \prod_{i=1}^N \frac{\prod_{j=1}^M x_{ij} [1 - P_{12}^{(i)}(j)]}{P^{(i)} \sum_{j=1}^M x_{ij} [1 - P_{12}^{(i)}(j)] + (1 - P^{(i)}) \sum_{j=1}^M x_{ij} P_{21}^{(i)}(j)} \geq D_{\text{г.з}}; \quad (8-11)$$

$$2) \sum_{j=1}^M x_{ij} = 1, \quad (i = 1, 2, \dots, N);$$

$$3) 1 - \frac{\prod_{i=1}^N P^{(i)} - \prod_{i=1}^N P^{(i)} A^{(i)}}{1 - \prod_{i=1}^N \left\{ P^{(i)} A^{(i)} + (1 - P^{(i)}) \sum_{j=1}^M x_{ij} P_{21}^{(i)}(j) \right\}} \geq D_{\text{нг.з}}; \quad (8-12)$$

$$A^{(i)} = \sum_{j=1}^M x_{ij} [1 - P_{12}^{(i)}(j)].$$

Учитывая, что $q^i = 1 - P^i \ll 1$; $P_{12}^{(i)}(j) \ll 1$; $P_{21}^{(i)}(j) \ll 1$, а также,

что $\prod_{i=1}^n (1 - z_i) \approx 1 - \sum_{i=1}^n z_i$ при $z_i \ll 1$, можно аппроксимировать ограничения (8-11), (8-12) более простыми, приближенными соотношениями, проведя над которыми дополнительные преобразования, получим следующую постановку задачи оптимизации для всего объекта контроля: найти

$$\left. \begin{aligned} & \min_{x_{ij}} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \bar{r}^{(i)}(j) \\ & \text{при ограничениях} \\ & 1) \sum_{i=1}^N q^{(i)} \sum_{j=1}^M x_{ij} P_{21}^{(i)}(j) \leq 1 - D_{\text{г.з}}; \\ & 2) \frac{\sum_{i=1}^N (1 - q^{(i)}) \sum_{j=1}^M x_{ij} P_{12}^{(i)}(j)}{\sum_{i=1}^N q^{(i)} \sum_{j=1}^M x_{ij} [1 - P_{21}^{(i)}(j)]} \leq \frac{1 - D_{\text{нг.з}}}{D_{\text{нг.з}}}; \\ & 3) \sum_{j=1}^M x_{ij} = 1, \quad i = 1, 2, \dots, N. \end{aligned} \right\} \quad (8-13)$$

Из (8-13) видно, что рассматриваемая задача является задачей дискретного программирования. Ниже приведены оценки целевой функции и функции ограничений, используемые в методе «ветвей и границ»:

$$T[X_1, 2, \dots, k(j_1, j_2, \dots, j_k)] =$$

$$= \sum_{i=1}^k \bar{r}^{(i)}(j_i) + \sum_{i=k+1}^N \min_{x_{ij}} \sum_{j=1}^M x_{ij} \bar{r}^{(i)}(j),$$

где $T[X_1, 2, \dots, k(j_1, j_2, \dots, j_k)]$ — оценка целевой функции для подмножества вариантов $X_1, 2, \dots, k(j_1, j_2, \dots, j_k)$, причем для k первых параметров алгоритмы принятия решений выбраны и равны $j_1, j_2, \dots, j_k$, а для остальных параметров $i = k + 1, k + 2, \dots, N$ выбор определенного способа достоверности не сделан. Оценка целевой функции принимается во внимание, если справедливы следующие оценки для ограничений:

$$\sum_{i=1}^k q^{(i)} P_{21}^{(i)}(j_i) + \sum_{i=k+1}^N q^{(i)} \min_{x_{ij}} \sum_{j=1}^M x_{ij} P_{21}^{(i)}(j) \leq 1 - D_{\text{г.з}}; \quad (8-14)$$

$$\frac{\sum_{i=1}^k (1 - q^{(i)}) P_{12}^{(i)}(j_i) + C}{\sum_{i=1}^N q^{(i)} (1 - P_{21}^{(i)}(j_i)) + D} \leq \frac{1 - D_{\text{нг.з}}}{D_{\text{нг.з}}}, \quad (8-15)$$

где

$$C = \sum_{i=k+1}^N (1 - q^{(i)}) \min_{x_{ij}} \sum_{j=1}^M x_{ij} P_{12}^{(i)}(j);$$

$$D = \sum_{i=k+1}^N q^{(i)} \max_{x_{ij}} \sum_{j=1}^M x_{ij} [1 - P_{21}^{(i)}(j)].$$

Принцип использования оценок (8-14), (8-15) при решении данной задачи оптимизации методом «ветвей и границ» остается таким же, как и в примере 8-1.

В заключение главы следует отметить, что в ряде случаев структурный синтез хотя и может быть формализован в виде оптимальных задач дискретного программирования, однако он трудно разрешим в силу большой размерности задач, наличия большого числа случайных факторов, из-за векторного характера показателя эффективности. В таких ситуациях структурный синтез обычно понимают как выбор структуры элементов и принципов их взаимодействия из некоторого небольшого числа вариантов, перспективных для дальнейшего использования. Предварительный отбор таких вариантов по совокупности свойств осуществляется группой квалифицированных специалистов, имеющих опыт проектирования подобных систем, а деталь-

ное исследование количественных показателей основных свойств этих вариантов проводится для каждого из них в отдельности. После этого проводится сравнительный анализ и окончательный выбор основного варианта построения структуры системы. В этом случае говорят, что структурный синтез ведется на основе методов анализа. При этом используются рассмотренные в предыдущих главах способы анализа различных функциональных свойств, широко применяемые на этапе макропроектирования систем (выбор структуры, алгоритмов и параметров взаимодействия элементов в системе).

СПИСОК ЛИТЕРАТУРЫ

1. Бешелев С. Д., Гурвич Ф. Г. Математико-статистические методы экспертных оценок. — М.: Статистика, 1974. — 160 с.
2. Бриллион Л. Научная неопределенность и информация. — М.: Мир, 1966. — 272 с.
3. Брук В. М., Николаев В. И. Начала общей теории систем. — Л.: СЗПИ, 1977. — 62 с.
4. Бусленко Н. П. Моделирование сложных систем. — М.: Наука, 1968.
5. Бусленко В. Н. Автоматизация имитационного моделирования сложных систем. — М.: Наука, 1977. — 240 с.
6. Бусленко Н. П., Калашников В. В., Коваленко И. Н. Лекции по теории сложных систем. — М.: Советское радио, 1973. — 438 с.
7. Вагнер Г. Основы исследования операций. — М.: Мир, 1972. — 332 с., т. 1.
8. Вентцель Е. С. Исследование операций. — М.: Советское радио, 1972. — 552 с.
9. Глушков В. М. Введение в АСУ. — Киев: Техніка, 1974. — 319 с.
10. Глушков В. М. Проектирование и внедрение АСУП. — Киев: Техніка, 1974. — 191 с.
11. Глушков В. М. Сети ЭВМ. — М.: Связь, 1977. — 279 с.
12. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. — М.: Наука, 1966. — 431 с.
13. Голенко Д. И. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах. — М.: Наука, 1972.
14. Гуд Г. Х., Маккол Р. Э. Системотехника. — М.: Советское радио, 1962.
15. Гурьянов А. И., Митрофанов Ю. И. Моделирование вычислительных систем разделения времени сетями систем массового обслуживания. — В кн.: Системное моделирование. — Новосибирск, 1971, вып. 2.
16. Денисов А. А. Основы теории информационных цепей. — Л.: ЛПИ, 1977. — 55 с.

17. Денисов А. А. Теоретические основы кибернетики. — Л.: ЛПИ, 1975. — 40 с.
18. Дехтяренко В. А., Своятыцкий Д. А. Методы многокритериальной оптимизации сложных систем при проектировании. — Киев: Техніка, 1976.
19. Дурандин К. П. Анализ эффективности мультипрограммной работы АСК в условиях разделения времени. — Тр. ЛПИ, 1974, № 337, с. 41—45.
20. Дружинин В. В., Конторов Д. С. Проблемы системологии. — М.: Советское радио, 1976. — 295 с.
21. Зайченко Ю. П. Исследование операций. — Киев: Высшая школа, 1975. — 319 с.
22. Захаров В. Н., Поспелов Д. А., Хазацкий В. Е. Системы управления. — М.: Энергия, 1977. — 424 с.
23. Ивашкин Ю. А. Структурный анализ и синтез человеко-машинных систем управления производством. — Приборы и системы управления, 1978, № 7, с. 4.
24. Ивинский В. А. Сети массового обслуживания и их применение в ЭВМ. — Зарубежная радиоэлектроника, 1977, № 7, с. 33—17.
25. Карлин С. Основы теории случайных процессов. — М.: Мир, 1971. — 536 с.
26. Коган Б. М., Корочатов А. Ф. Вероятностные модели информационно-вычислительных сетей на основе динамики средних. — Автоматика и телемеханика, 1975, № 3, с. 100—110.
27. Колмогоров А. Н. Три подхода к определению понятия количества информации. — В кн.: Проблемы передачи информации. — М.: Наука, 1965, т. 1, с. 4—13.
28. Кофман А. Введение в прикладную комбинаторику. — М.: Наука, 1975. — 480 с.
29. Кофман А., Анри-Лабурдер А. Методы и модели исследования операций. — М.: Мир, 1977. — 432 с.
30. Кофман А., Крюон Р. Массовое обслуживание, теория и применение. — М.: Мир, 1965.
31. Кудрицкий М. А., Сеница М. А., Чинаев П. И. Автоматизация контроля РЭА. — М.: Советское радио, 1977. — 254 с.
32. Ломов Б. Ф. Человек и техника. — М.: Советское радио, 1966. — 464 с.
33. Мартин Ф. Моделирование на вычислительных машинах. — М.: Советское радио, 1972. — 277 с.
34. Месарович М. Д., Мако Д., Такахара И. Теория иерархических многоуровневых систем. — М.: Мир, 1973. — 344 с.
35. Нейлор Т. Машинные имитационные эксперименты с моделями экономических систем. — М.: Мир, 1975. — 500 с.
36. Нейман Л. Р., Демирчян К. С. Теоретические основы электротехники. — М.: Энергия, 1966. — 522 с.
37. Нечипоренко В. И. Структурный анализ систем. — М.: Советское радио, 1977. — 216 с.
38. Петров Б. Н. Начала информационной теории управления. — Техническая кибернетика. — М.: ВИНТИ, 1971. — 240 с., ч. 3.
39. Поспелов Д. А. Большие системы. Ситуационное управление. — М.: Знание, 1975. — 64 с.
40. Поспелов Г. С. Программно-целевое планирование и управление. — М.: Советское радио, 1976. — 440 с.
41. Поляк Ю. Г. Вероятностное моделирование на электронных вычислительных машинах. — М.: Советское радио, 1971. — 400 с.

42. Розенберг В. Л., Прохоров А. И. Что такое теория массового обслуживания? — М.: Советское радио, 1965. — 255 с.
43. Саати Т. Л. Элементы теории массового обслуживания и ее приложения. — М.: Советское радио, 1971. — 520 с.
44. Садовников В. И., Энштейн В. Л. Потоки информации в системах управления. — М.: Энергия, 1974. — 240 с.
45. Соболев И. М. Численные методы Монте-Карло. — М.: Наука, 1973.
46. Советов Б. Я. Основы построения АСУ. — Л., Изд-во ЛГУ, 1975. — 132 с.
47. Темников Ф. Е. Теоретические основы информационной техники. — М.: Энергия, 1971. — 424 с.
48. Федоренко Н. П. Математика и кибернетика в экономике. — М.: Экономика, 1976. — 700 с.
49. Федоренко Н. П. Математические методы в экономических исследованиях. — М.: Наука, 1974. — 172 с.
50. Хазацкий В. Е. Управляющие машины и системы. — М.: Энергия, 1976. — 246 с.
51. Холл А. Д. Опыт методологии для системотехники. — М.: Советское радио, 1975. — 446 с.
52. Цвирун А. Д. Структура сложных систем. — М.: Советское радио, 1975. — 198 с.
53. Шапот Д. В. О построении критериев качества технических объектов. — Изв. АН СССР. Сер. техн. кибернетика, 1971, № 6, с. 58—59.
54. Шеннон К. Работы по теории информации и кибернетике. — М.: Иностран. лит., 1963. — 829 с.
55. Шнепс М. А. Численные методы теории телетрафика. — М.: Связь, 1974. — 232 с.
56. Яковлев Е. И. Машинная имитация. — М.: Наука, 1975. — 157 с.
57. Янбых Г. Ф., Этингер Г. Ф. Проектирование структуры отраслевой сети вычислительных центров. — Л.: Энергия, 1974. — 104 с.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Автоматизированные системы управления 7
Агрегат 73
Агрегативное представление систем 61
Аддитивное преобразование 29
Альтернатива 2

Бесконечный источник 102
Большая система управления 5

Граф неориентированный 43
— несвязный 48
— ориентированный 43
— связный 48

Дерево решений 269
Децентрализация 77
Дисциплина обслуживания 104

Замкнутая сеть массового обслуживания 122

Иерархия 32, 228
Иерархическая структура 32
Избыточность 230
Имитация 146
Информация 174
Инцидентность 44
Источник
— конечный 102

Квазиравномерность случайных чисел 160
Контур 66
Координация 86
Критерий эффективности 27

Макропроектирование 20
Марковский процесс 91

Математическое программирование 28
Метод «ветвей и границ» 268
— динамики средних 141
— обратной функции 164
— экспертных оценок 32
Модель 17
Мощность смысловая 192
Мультипликативное преобразование 29
Мультипрограммный режим 117

Нагрузка 178

Обмен информацией 7
Объем имитационных экспериментов 170
Отражение 216
Оценка 171

Память 196
Подграф 46
Пополнение численности 145
Порядковая функция на графе 48
Потенциал 176
Поток 182
Приемник 182
Принцип квазирегулярности 143
Приоритет 104
Процесс «гибель и размножение» 108
Псевдослучайность 161
Путь 46

Разомкнутая сеть 120
Ранг 59
— критерия 33
Рассредоточенная структура 78
Ригидность 201

Связность 48	Управление 77
Система массового обслуживания 101	Уровень графа 48
—с дискретными событиями 151	
Системный подход 17	Целевая функция 32
Системотехника 17	Централизация 58
Смежность 43	Цепь 46
Смысл 192	Цикл 46
Состояние 68	
Стохастическая сеть 120	Частичный граф 46
Структура 39	Числовая функция на графе 49
Структурный синтез 267	
Сущность 229	Эффективность 32
Тест пар 163	Язык моделирования 173
— частот 162	
Ток информационный 179	

ОГЛАВЛЕНИЕ

Предисловие	3
Глава первая. Системный подход к изучению больших систем управления	5
1-1. Понятие БСУ	—
1-2. Системный подход. Основные задачи системотехники	17
1-3. Принципы построения обобщенных критериев	27
Глава вторая. Структурный анализ больших систем управления	39
2-1. Цели и задачи структурного анализа	—
2-2. Формализация описания структуры на основе теории графов	42
2-3. Описание и анализ потоков информации в БСУ	51
2-4. Топологическая декомпозиция структур	54
2-5. Структурно-топологические характеристики систем и их применение	56
2-6. Модель структурного сопряжения элементов в БСУ	61
Глава третья. Агрегативные модели функционирования больших систем управления	67
3-1. Общая модель функционирования	—
3-2. Агрегативные модели элемента и системы	73
3-3. Управление в больших системах	77
Глава четвертая. Применение марковских процессов для анализа поведения больших систем управления	91

4-1. Марковские процессы с конечным или счетным множеством состояний и непрерывным временем	91
4-2. Представление БСУ в виде моделей систем массового обслуживания и их классификация	101
4-3. Модели систем массового обслуживания, описываемые процессами типа «гибель и размножение», и их характеристики	108
Глава пятая. Методы анализа поведения систем при большом числе элементов	120
5-1. Представление систем в виде стохастических сетей. Разомкнутые сети массового обслуживания	—
5-2. Замкнутые сети массового обслуживания и анализ их характеристик	127
5-3. Метод динамики средних	141
Глава шестая. Применение имитационного моделирования для исследования характеристик больших систем управления	146
6-1. Понятие об имитационном моделировании. Основные его этапы	—
6-2. Имитация функционирования систем с дискретными событиями	151
6-3. Методы имитации случайных факторов при программном моделировании систем	160
6-4. Имитация функционирования БСУ, описываемых в классе марковских процессов	168
6-5. Определение объема имитационных экспериментов	170
Глава седьмая. Информационный подход к анализу больших систем управления	174
7-1. Информационные цепи. Установившийся режим	—
7-2. Переходные режимы в информационных цепях	196
7-3. Методология системного анализа	216
7-4. Анализ иерархических структур	228
7-5. Управление системами подвижных объектов	245
Глава восьмая. Структурный синтез больших систем управления	254
8-1. Агрегатизация—основа построения структуры современных БСУ	—
8-2. Постановка задачи структурного синтеза	263
8-3. Основные принципы построения алгоритмов структурной оптимизации	267
Список литературы	282
Предметный указатель	285

**АНАТОЛИЙ АЛЕКСЕЕВИЧ ДЕНИСОВ
ДМИТРИЙ НИКОЛАЕВИЧ КОЛЕСНИКОВ
ТЕОРИЯ БОЛЬШИХ СИСТЕМ УПРАВЛЕНИЯ**

Редактор В. А. Романов
Художественный редактор Д. Р. Степанович
Технический редактор А. Г. Рябкина
Корректор А. Н. Акимов
Переплет художника В. В. Костырева

ИБ № 1916

Сдано в набор 11.08.81. Подписано в печать 01.02.82. М-21826.
Формат 84×108¹/₃₂. Бумага типографская № 2. Гарнитура литературная. Печать высокая. Усл. печ. л. 15,12. Уч.-изд. 17,13.
л. Тираж 20 000 экз. Заказ 47. Цена 85 к.

Ленинградское отделение Энергоиздата. 191041, Ленинград,
Марсово поле, 1.

Набрано и сматрицировано: ордена Октябрьской Революции,
ордена Трудового Красного Знамени Ленинградском производственно-техническом объединении «Печатный Двор» имени А. М. Горького Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 197136, Ленинград, П-136, Чкаловский пр., 15.

Отпечатано во Владимирской типографии «Союзполиграфпрома» при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 600000, г. Владимир, Октябрьский пр., 7.