И. А. УШАКОВ

ВЕРОЯТНОСТНЫЕ МОДЕЛИ НАДЕЖНОСТИ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

МОСКВА „РАДИО И СВЯЗЬ"

УДК 681.324+519.248

Ушаков И. А. Вероятностные модели надежности информационно-вычислительных систем. — М.: Радио и связь, 1991. — 132 с.: ил. — ISBN 5-256-00795-5.

Рассматриваются математические модели, используемые при априорных вероятностных расчетах, апостериорных статистических оценках и оптимизации показателей надежности информационно-вычислительных систем (ИВС).

Дается описание абстрактного процесса функционирования, приводятся основные показатели надежности ИВС и обсуждается проблема их выбора. Наряду с известными вопросами теории надежности рассматриваются методы расчета надежности резервированных систем с учетом контроля и переключения, а также расчеты вероятности выполнения вычислительных задач на ЭВМ с учетом простоев и сбоев.

Излагаются задачи оптимального резервирования, рассматривается вопрос об оптимальной периодичности контроля. Приводится новый метод оптимизации на основе обработки статистических реализаций, полученных при имитационном моделировании.

Материал иллюстрируется примерами применительно к ИВС и средствам вычислительной техники.

Для научных работников, может быть полезна инженерам и математикам, занимающимся обеспечением надежности вычислительных систем.

Табл. 16. Ил. 20. Библ. 29 назв.

Рецензент: академик АН УССР *Б. В. Гнеденко*

Редакция литературы по информатике и вычислительной технике

Научное издание

Ушаков Игорь Алексеевич

ВЕРОЯТНОСТНЫЕ МОДЕЛИ НАДЕЖНОСТИ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Заведующая редакцией Г. И. Козырева
Редактор Т. М. Бердичевская
Обложка художника Н. А. Пашуро
Художественный редактор Н. С. Шеин
Технический редактор А. Н. Золотарева
Корректор Т. Л. Кускова

ИБ № 2331

Сдано в набор 12.12.90. Подписано в печать 27.03.91
Формат 60×90/16. Бумага типогр. № 2. Гарнитура литературная
Печать высокая. Усл. печ. л. 8,25. Усл. кр.-отт. 8,63
Уч.-изд. л. 9,09. Тираж 3000 экз. Изд. № 23280
Зак. № 1205. Цена 2 р.
Издательство «Радио и связь». 101000 Москва, Почтамт, а/я 693

Типография издательства «Радио и связь». 101000 Москва, Почтамт, а/я 693

У $\frac{2404000000-072}{046(01)-91}$ 23-91

ISBN 5-256-00795-5

ПРЕДИСЛОВИЕ

Проблема надежности, возникшая много лет назад, заставила говорить о ней в первую очередь инженеров и экономистов, а затем и математиков как о проблеме номер 1. Вначале речь шла не о низкой надежности технических средств в настоящем смысле слова. Безотказность и даже ремонтопригодность техники были вполне удовлетворительными, но затраты на восстановительные работы, замену отказавших деталей и различные технические профилактические мероприятия были слишком велики. По известным источникам[1], в 1949 г. около 70 % всей морской радиоэлектронной аппаратуры США находилось в состоянии ремонта. В конце второй мировой войны около 60 % самолетного оборудования, переброшенного на Дальний Восток, оказалось неисправным, при этом около 50 % запасных комплектов и элементов вышли из строя в результате хранения. В тот период радиосвязное оборудование находилось в неработоспособном состоянии 1/7 часть всего времени эксплуатации, радиолокационное — 5/6, гидроакустическое — около 1/2 времени. Перечисление можно было бы продолжить для других видов техники, других стран и других периодов времени.

Прошло чуть более четверти века, и мир заговорил о научно-технической революции. Начали создаваться сверхсложные системы в информатике, энергетике, транспорте и в других отраслях народного хозяйства. Причем это были не просто системы, которые характеризовались большим числом входящих в их состав элементов, сложными структурой и алгоритмами функционирования. Это были системы, пронизывающие всю инфраструктуру современного общества на государственном уровне, а это приводило не только к чисто структурному и функциональному их усложнению, но и резкому повышению требований к надежности, живучести и безопасности функционирования. Этот период развития техники характеризуется уже не только лозунгами о важности проблемы надежности, но и бурным развитием методов обеспече-

[1] Надежность наземного радиоэлектронного оборудования: Пер. с англ. / Под ред. Н. М. Шулейкина. — М.: Сов. радио, 1957.

ния высокой надежности систем на всех этапах: при проектировании, производстве, испытаниях и эксплуатации.

Действительно, проектирование и реализация сложных технических систем, на создание которых в течение многих лет затрачивались огромные людские и материальные ресурсы, уже невозможно было осуществлять «на глазок». Требовался строгий математический расчет всех технических параметров, включая различные показатели надежности, нужны были обоснованные технико-экономические решения. При этом, учитывая огромную ответственность задач, решаемых техническими сверхсистемами на уровне национальной экономики, национальной безопасности, порою непредвидимые экономические и морально-политические последствия от возможных ошибок и отказов в этих системах, необходимо было не только обеспечить технические возможности этих систем вообще, но и, что самое главное, сохранить и поддержать работоспособность этих систем в течение очень длительного времени эксплуатации.

Можно привести примеры многих современных технических систем, для которых решение проблемы надежности в самом прямом смысле означает, быть или не быть данной системе. К ним можно отнести различные системы информатики: региональные и отраслевые автоматизированные системы управления, включающие в свой состав большое число ЭВМ, системы управления воздушным движением для гражданской авиации, автоматизированные системы управления технологическими процессами, сети центров управления и слежения за космическими объектами, сети и системы передачи данных.

Усложнение систем идет в различных направлениях. С одной стороны, в состав технических систем входит все большее число комплектующих элементов. С другой стороны, усложняется их структура, определяющая соединение отдельных элементов и их взаимодействие в процессе функционирования и поддержания работоспособности. Понятно, что усложнение систем является прямым следствием постоянно возрастающей ответственности выполняемых ими функций, сложности и многообразия этих функций, что, в свою очередь, диктуется прогрессом науки и техники.

При прочих равных условиях система, состоящая из большого числа комплектующих элементов и имеющая более сложную структуру и сложный алгоритм функционирования, является менее надежной по сравнению с более простой системой. Это требует разработки специальных методов обеспечения, повышения и поддержания надежности таких систем, включая разработку математических методов априорных расчетов и экспериментальной оценки.

Инженеры, физики и математики приложили немало совместных усилий для разработки современной теории надежности. Были предприняты гигантские усилия для создания более надежных компонентов, более простых и надежных схем и конструкций, улучшения условий эксплуатации. Были разработаны соответствующие методы, позволяющие осуществлять анализ и синтез разрабатываемых технических средств на этапе проектирования, проводить обоснованные оценки показателей надежности этих средств во время испытаний и эксплуатации.

Однако проблема надежности продолжает оставаться одной из основных для современной техники. Дело, видимо, объясняется не столько тем, что достигнутая надежность современных технических систем слишком низка, сколько тем, что непрерывно усложняются решаемые задачи и одновременно повышаются требования к надежности их выполнения.

Прежде чем перейти к анализу роли математических методов теории надежности в решении проблемы обеспечения высокой надежности функционирования современных сложных технических систем, отметим, что развитие математической теории надежности — обязательный атрибут современной технологии создания этих систем: выдача количественных рекомендаций (в терминах задач надежности — это априорные вероятностные расчеты на ранних стадиях создания, решение различных оптимизационных задач на этапах разработки и эксплуатации, проведение апостериорных статистических оценок по результатам специальных испытаний и реальной эксплуатации) — обязательное условие обоснованных и грамотных с технико-экономической точки зрения инженерных решений.

1. МЕТОДИЧЕСКИЕ ВОПРОСЫ НАДЕЖНОСТИ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

1.1. О построении математических моделей

Основная задача теории надежности — создание математических моделей реальных технических объектов, процессов их функционирования и технического обслуживания, а затем разработка математических методов исследования этих моделей с целью получения количественных характеристик. Без проведения количественного анализа невозможна разработка конкретных и эффективных мероприятий по обеспечению высокой надежности технических систем.

Особенностью математических моделей надежности информационно-вычислительных систем (ИВС) является то, что эти модели имеют дело с вероятностными процессами и используют в качестве исходных данных достаточно недостоверную статистику, иногда эта статистика вообще отсутствует (первое относится, как правило, к данным по надежности, а второе — к информации о потоках в сетях связи, о загрузке ЭВМ, об интенсивности межабонентской связи и т. п.). Тем не менее создание математических моделей в конечном счете направлено на количественное оценивание уровня надежности и эффективности функционирования систем информатики. Модель позволяет формировать различные рекомендации по повышению надежности за счет правильного и обоснованного выбора структуры системы, правил ее использования, методов технического обслуживания. Если не всегда можно с полным основанием доверять абсолютным значениям вычисленных априорно или оцененных статистически тех или иных показателей надежности, то для сравнительного анализа надежности различных вариантов построения или использования технических систем математические методы расчета надежности являются поистине незаменимыми.

Чем сложнее система и принцип ее организации (структура соединения элементов, взаимосвязь при функционировании, характер технического обслуживания и материального обеспечения и т. д.), тем эффективнее на любом этапе проектирования, раз-

работки и эксплуатации использование математических методов анализа и синтеза.

Каждая математическая модель является, безусловно, лишь грубым абстрактным описанием (и то лишь в той или иной плоскости) реального объекта. Она отражает лишь те стороны реального явления, которые являются решающим в данном конкретном исследовании. Другие, менее существенные свойства явления могут быть полностью проигнорированы. Это может привести к тому, что при более глубоком и всестороннем исследовании реального явления могут оказаться недостаточными взятые в исходной модели характеристики, параметры, свойства: модель может оказаться в определенном смысле несостоятельной. Не зря физики шутят, что «самой лучшей моделью кошки может быть кошка, а идеальной моделью любой кошки может быть только та же самая кошка».

Естественно, что та или иная математическая модель отображает степень нашего познания исследуемой технической системы. К сожалению, опыт показывает, что априорные представления даже о сравнительно общих принципах функционирования создаваемых сложных систем иногда весьма далеки от истины. В особенности это относится к различным системам информатики. Так, в сложных сетях передачи данных протоколы сетевого уровня часто бывают столь сложными в смысле описания, что допускают лишь имитационные способы математического моделирования; отсутствие в настоящее время хороших критериев степени верификации программного обеспечения вычислительных систем не позволяет создавать адекватных математических моделей вычислительных процессов и т. д. Однако необходимость исследования созданной системы в целях устранения различных неполадок, отыскания путей ее улучшения, разрабтки методов рациональной эксплуатации и т. д. приводит к необходимости более глубокого изучения системы.

В этом смысле любая «наилучшая» математическая модель процессов функционирования сложной системы является лишь наиболее полным возможным приближением к исследуемому процессу на данном уровне его понимания. Уточнение математической модели возможно лишь при дальнейшем изучении реального объекта, при сравнении теоретических результатов с опытными данными. Процесс создания адекватной математической модели в теории надежности заключается не только в теоретической разработке какой-либо гипотезы о реальном поведении объекта, но и в постоянной проверке соответствия принятой гипотезы имеющимся статистическим данным, получаемым опытным путем.

Итак, более глубокое исследование системы позволяет строить модель, точнее соответствующую реальной системе. Но более сложная матматическая модель требует, как правило, более детальных исходных данных, с одной стороны, и более тонких методов математического исследования — с другой. И хотя, казалось бы, подобное уточнение математической модели является желательным и даже необходимым для более точного изучения исследуемого объекта, возникает далеко непраздный вопрос: нужно ли стремиться к тому, чтобы математическая модель надежности системы была очень точной?

Чрезмереное уточнение математической модели не всегда имеет смысл. Дело в том, что сама по себе математическая модель решает далеко не все. Как правило, для получения количественных результатов мы пользуемся исходными данными, получаемыми экспериментальным путем на основании достаточно ограниченного числа опытных данных, т. е. не являющихся достаточно достоверными. Кроме того, если математическая модель надежности сложна, приходится прибегать к различным вычислительным методам, приводящим к неизбежным погрешностям (например, численным методам приближенных вычислений, асимптотическим методам, статистическим вычислительным методам и пр.). Эти два фактора — недостоверность (или неточность) исходных данных и погрешности вычислительных методов — могут свести на нет все те преимущества, которых мы пытаемся добиться, создавая очень точную математическую модель. Естественно, возникает вопрос о целесообразности создания точной математической модели исследуемой системы. Иными словами, сама по себе «чистая» модель надежности не является полностью определяющим средством исследования реальной системы и точность ее должна определяться конкретными условиями: требуемой точностью исследования, достоверностью различных исходных данных, возможной точностью численных расчетов и т. д.

Малая достоверность исходных статистических данных, неточность математической модели (невозможность учета всех факторов, идеализация отдельных процессов и т. д.) и, как следствие этого, погрешности в окончательных результатах могут зародить сомнение в полезности расчетов надежности. Поэтому крайне важно понять, когда и для чего нужны расчеты надежности.

Во-первых, расчеты надежности функционирования, безусловно, приносят большую пользу на ранних этапах проектирования, когда возникает вопрос о сравнении различных возможных вариантов построения системы, ее архитектуры, организации вычислительных и информационных процессов и выборе наилучших из них.

Во-вторых, расчеты надежности производятся на стадии технического проектирования, когда уже более детально известны состав системы, ее структура и принципы функционирования, позволяют проверить правильность принятых решений, найти слабые места и выработать определенные рекомендации по повышению надежности эффективности функционирования.

В-третьих, расчетные методы часто оказываются незаменимыми, а порой и единствено возможными на этапе испытания сложных систем информатики. Часто очень большие и сложные системы приходится испытывать либо по частям (причем обычно в течение разного времени), либо даже не в полном штатном составе. В обоих случаях единственным способом получить оценку показателей надежности является расчетно-экспериментальный метод.

В-четвертых, только расчетные и расчетно-экспериментальные методы расчета надежности дают возможность обоснованно планировать и прогнозировать стратегию модернизации и развития больших систем информатики: систем передачи данных, вычислительных сетей и т. п.

В-пятых, именно расчетные методы (по организации контроля исправности, по проведению профилактического обслуживания и т. п.) могут обеспечить рациональный режим эксплуатации средств информатики.

Наконец, составление математических моделей структуры взаимодействия элементов системы, процессов функционирования, правил технического обслуживания приводит к необходимости все глубже вникать в суть создаваемой технической системы, все больше понимать ее специфические особенности и тем самым не только количественно оценивать характеристики надежности, но и вырабатывать «попутно» полезные качественные рекомендации по повышению надежности и обеспечению оптимальных режимов функционирования и обслуживания.

Следует подчеркнуть, что чем сложнее исследуемая система, тем эффективнее использование математических расчетных методов на всех этапах ее разработки и использования.

1.2. Математические модели надежности систем информатики

Системы информатики, как уже отмечалось, представляют собой сложные территориально распределенные системы, в состав которых входят тысячи комплектующих элементов. Однако эти системы характеризуются не только большим числом элементов — чисто количественная сторона естественным образом перерастает в новое качество: таким системам присущи сложная структура, сложный принцип функционирования, новые характерные особенности процесса эксплуатации. Кроме того, эти системы, как правило, являются развивающимися, т. е. в процессе своего существования они постоянно модернизируются, меняются условия их работы и выполняемые задачи, возникают различные проблемы, которых не было на ранних стадиях существования этих систем.

Понятно, что разработка математических моделей, адекватно описывающих процесс функционирования таких систем, оказывается весьма сложной и трудоемкой. Следует сразу же напомнить, что любая математическая модель отражает лишь часть определенных свойств исследуемой системы, причем в меру наших знаний самой исследуемой системы. В определенном смысле моделируется не сама система, как нам хотелось бы, а наше представление о ней. Тем не менее наши даже порой чисто качественные и умозрительные представления о структуре системы с учетом всех существующих в ней связей, о характере ее функционирования с учетом всей возможной неопределенности различных внешних и внутренних факторов могут привести к построению таких математических моделей, анализ которых окажется невозможным даже при использовании самых современных ЭВМ: эти модели будут слишком сложны, слишком большой размерности, исходные данные о различных важных факторах будут либо весьма недостоверными, либо вообще неизвестными.

Из этих замечаний следует, что к построению математических моделей необходимо относиться как к искусству: нужно уметь строить достаточно простую модель, которая при этом могла бы все же приводить к конструктивным выводам об исследуемой системе (хотя бы на качественном уровне).

Какие общие требования могут быть сформулированы для специалиста, занимающегося построением математических моделей для таких сложных технических объектов, какими являются системы информатики?

1. Математическая модель должна отражать основные свойства исследуемого объекта с точки зрения интересующего параметра или группы параметров. Например, если рассматривается время доставки сообщений в системе с пакетной коммутацией, то подходящей математической моделью может быть сеть массового обслуживания с ненадежными обслуживающими устройствами (узлами коммутации, каналами связи). Если же изучается вопрос живучести сети по отношению к различного рода внешним воздействиям, то при выборе в качестве основного показателя вероятности связности подходящей математической моделью может оказаться граф с сетевой структурой.

2. Математическая модель должна быть достаточно простой в содержательном смысле, т. е. результаты ее анализа должны быть легко интерпретируемы. Это означает, например, что слишком подробная модель, обеспечивающая одновременное получение большого числа взаимосвязанных параметров, может и не быть наилучшим вариантом. Например, получение совместного распределения числа обслуживаемых заявок в сети связи в различных узлах коммутации может оказаться той информацией, которую исследователь не умеет принципиально использовать. Однако эта модель позволяет анализировать «узкие» места в сети связи, в которой используются гораздо меньшие и более обозримые объемы выходной информации.

3. Математическая модель должна быть «адаптированной» под имеющиеся исходные данные. Например, бесмысленно строить полумарковскую модель для описания процесса функционирования системы с восстановлением, если о распределениях известны лишь значения математических ожиданий. Если при таких же исходных данных требуется оценить значения структурных параметров сети, то нужно иметь в виду, что могут быть получены лишь двусторонние оценки, а не точные значения соответствующих параметров.

Как правило, наличие неполных или слишком недостоверных данных (а именно это и является, к сожалению, отличительной особенностью данных о надежности) делает очень оправданным упрощение математических моделей.

4. Математическая модель должна быть легко модифицируемой при появлении новых исходных данных или новых сведений о внутренней природе системы. Например, статистическая модель системы передачи данных в режиме пакетной коммутации должна предусматривать возможность использования ее при различных протоколах верхнего уровня. Математическая модель для статистического моделирования системы связи с коммутацией сообщений должна иметь общее программное ядро с математической моделью системы связи с коммутацией каналов.

5. Наконец, математическая модель системы информатики — объекта весьма сложного и содержащего огромное число входящих в ее состав подсистем и устройств — должна быть сформулирована так, чтобы размерность этой модели позволяла бы проводить достаточно конструктивным образом расчеты на доступной вычислительной технике в разумные сроки. В первую очередь это касается тех моделей, которые предназначены для использования в режиме советчика в реальном масштабе времени.

1.3. Основные термины и определения

Прежде чем начать описание математических моделей надежности, введем основные термины и определения, принятые в современной инженерной практике.

Надежностью технической системы называется свойство системы сохранять работоспособность в заданных условиях функционирования. Говоря о работоспособности, следует сразу же определить критерий отказа системы. Эти два понятия в определенном смысле выражаются одно через другое: отказ — это потеря работоспособности. Однако для данной технической системы конкретное определние отказа зависит от многих факторов: назначения системы, выполняемой задачи, требований к выполнению данной конкретной функции. Отметим также, что возможны не только нарушения работоспособности, но и нарушение функционирования системы без потери работоспособности. Действительно, для системы связи и передачи данных могут возникать отказы по критериям, важным для пользователей (чрезмерные задержки в передаче информации, потери сообщений и т. п.) даже при полной исправности всех элементов системы из-за случайного роста загрузки или из-за нерациональных алгоритмов функционирования (например, из-за несовершенства протоколов сетевого уровня в системах передачи данных). Такие отказы будут рассмотрены далее. Вообще под надежностью будем понимать только внутреннее свойство, присущее данной системе.

Надежность — это сложное свойство, включающее в свой состав несколько единичных свойств: безотказность, готовность, сохраняемость, ремонтопригодность. Иногда в это свойство включают также понятия безопасности и живучести.

Под безопасностью технической системы понимают ее способность функционировать в основном режиме, а также находиться в состоянии аварийного или планового ремонта, не причиняя вреда здоровью обслуживающего персонала и окружающей среде. Для систем информатики это свойство не является существенным по сравнению, например, с системами энергетики, и в первую очередь с объектами атомной энергетики [1].

Под живучестью технической системы понимают ее способность противостоять крупномасштабным внешним воздействиям как естественного характера (природные стихийные бедствия, неблагоприятные погодные условия и т. п.), так и преднамеренным.

Как правило, при анализе живучести системы воздействиям, приводящим к ее разрушению, не приписывается вероятностный смысл и они рассматриваются без развития во времени. В этом случае вместо вероятностного подхода естественным оказывается минимаксный [1—4].

Отличительным признаком надежности как свойства технической системы является то, что она характеризуется вероятностными процессами, протекающими во времени. Дело в том, что надежность определяется изменением внутреннего состояния устройств системы во времени под воздействием окружающей среды (температура, влажность, вибрация) и внутренних физико-химических процессов, которые сами по себе имеют стохастическую природу.

*Событие, состоящее в полной или частичной утрате работо*способности, называется отказом. В этом определении обнаруживается некоторая противоречивость и неоднозначность. Во-первых, как только мы говорим о частичной утрате работоспособности, сразу же возникает вопрос: где грань, которая разделяет допустимый уровень ухудшения характеристик от недопустимого? Во-вторых, может быть не ясно, что такое полный отказ. Ведь, строго говоря, все процессы, связанные с отказами, являются непрерывными во времени, а потому объект, отказывая «полностью», обязательно проходит все фазы частичных отказов. Отсюда видно, как важна для анализа надежности технических систем четкая формулировка критерия отказа.

Скорость протекания процессов, сопутствующих возникновению и проявлению отказов, обусловливает деление их на внезапные и постепенные. Суть деления в том, что нам не удается наблюдать процесс постепенного изменения параметров, и поэтому многие постепенные отказы мы замечаем лишь тогда, когда они привели к весьма заметным последствиям. Естественно, что, узнавая больше о характере физико-химических процессов, производя более частые и более точные контрольные замеры различных параметров, мы можем все с большей точностью предсказывать те или иные события.

В теории надежности принято рассматривать устройства двух уровней сложности:

1) некоторое простейшее (в пределах данного конкретного исследования) устройство, которое удобно назвать элементом. Элемент — это такой объект, отдельные части которого не представляют существенного интереса в пределах проводимого анализа;

2) система — это определенная совокупность элементов, взаимодействующих в процессе выполнения рассматриваемого круга задач и взаимосвязанных функционально.

Относительность понятий «элемент» и «система» понятна. Подразделение системы на элементы зависит от требуемой точности проводимого анализа, от уровня наших представлений о системе, наконец, от квалификации и даже технических и научных «вкусов» исследователя. Более того, объект, считавшийся системой в одном исследовании, может рассматриваться как элемент, если изучается какая-либо система большего масштаба.

Например, в вычислительной сети элементом может считаться ЭВМ, терминал, канал связи. В то же время, рассматривая функционирование ЭВМ, можно выделить процессор, входные и выходные устройства, различные интерфейсы и т. д. Рассматривая устройства памяти, можно в качестве элементов выделить, например, дисководы, магнитные носители, элементы записи и считывания информации и т. п. Степень агрегированности элемента в каждом конкретном случае определяется целью исследования, характером выбранного показателя надежности и другими факторами.

В теории надежности весьма важную роль играет деление элементов и систем на восстанавливаемые и невосстанавливаемые. Содержательный смысл этих понятий очевиден. Следует, может быть, отметить, что часто встречаются устройства, которые в одни определенные периоды времени являются восстанавливаемыми, а в другие — невосстанавливаемыми. Так, устройство может быть восстанавливаемым в режиме дежурства (ожидания начала выполнения задачи), но в то же время является невосстанавливаемым при выполнении определенной задачи из-за невозможности каких-либо перерывов самого технологического процесса. Например, механическая деформация магнитного диска может сделать его полностью неработоспособным, причем каких-либо способов восстановления его может и не существовать. В этом смысле диск является невосстанавливаемым по отношению к отказу данного типа. Однако вычислительная система может быть вновь приведена в состояние работоспособности, если отказавший диск будет заменен на работоспособный.

Можно привести и более сложный пример, когда одно и то же устройство может считаться восстанавливаемым или невосстанавливаемым в зависимости от вида выполняемой задачи. Так, ЭВМ, работающая в режиме пакетной обработки вычислительных задач, является восстанавливаемой системой, так как после любого возникшего отказа она через некоторое время может продолжить выполнение своих функций. Та же ЭВМ, работающая в реальном масштабе времени например, управляя каким-либо непрерывным технологическим процессом), на время выполнения задачи являет-ся с точки зрения надежности невосстанавливаемой, так как любой ее отказ приводит к нарушению данного технологического процесса и восстановление ЭВМ без ущерба для этого процесса принципиально невозможно.

1.4. Описание абстрактного процесса функционирования

Если отвлечься от конкретного содержания процесса функционирования системы, то в простейшем случае можно предложить следующую его математическую модель. В любой произвольный момент времени каждый элемент системы может находиться в одном из двух состояний: отказа и работоспособности. Обозначим неизвестное текущее состояние i-го элемента ($i=1, ..., n$) в момент времени t через $s_i(t)$, состояние отказа через $\bar{s}_i$ и состояние работоспособности через s_i. Весь процесс функционирования элемента можно представить чередующейся последовательностью случайных величин $\xi_1^i, \eta_1^i, \xi_2^i, \eta_2^i, \ldots, \xi_j^i, \eta_j^i$, где ξ_j^i — длительность j-го периода работоспособности; η_j^i — длительность j-го периода отказа i-го элемента (в течение этого времени производятся восстановительные работы, если они в принципе возможны), т. е.

$$s_i(t)=\begin{cases} s_i, & \text{если} \quad t\in\xi_j^i, \\ \bar{s}_i, & \text{если} \quad t\in\eta_j^i, \quad j=1, 2, \ldots . \end{cases}$$

Наиболее часто используемой математической моделью такого случайного процесса является альтернирующий регенерирующий процесс, или альтернирующий процесс восстановления.

Такой процесс представляет собой чередующуюся последовательность взаимно независимых случайных величин: нечетные по номеру случайные величины имеют функцию распределения $F(t)=\mathscr{P}\{\xi\leqslant t\}$, а четные — $G(t)=\mathscr{P}\{\eta\leqslant t\}$. Конечно, независимость случайных интервалов ξ и η, а также интервалов $\xi_1, \xi_2, ..., (\eta_1, \eta_2, ...)$ между собой является определенной математической идеализацией реального процесса функционирования технической системы. Действительно, если, например, имеет место процесс накопления повреждений в системе, состоящей из большого числа составляющих элементов, то естественно ожидать, что после длительного периода работоспособности ξ потребуется и большее в среднем время восстановления η. Далее, если система состоит из стареющих во времени элементов, то следует предполагать и то, что случайные величины ξ_k в среднем будут уменьшаться с ростом номера k, если производится ремонт, а не полная замена отказавшего элемента.

Однако при инженерных расчетах в качестве математической модели процесса функционирования обычно используется именно

альтернирующий процесс восстановления, поскольку, с одной стороны, он достаточно адекватно описывает реальную картину, а с другой, его математическое исследование достаточно просто и результативно. Кроме того, существующие статистические исходные данные таковы, что не позволяют использовать их для расчета показателей надежности на основании математических моделей, учитывающих описанные выше формы зависимости случайных величин. Для получения соответствующих исходных статистических данных могут потребоваться либо специальные испытания, либо специальным образом организованная регистрация отказов при эксплуатации.

Рассмотрим систему, состоящую из n элементов. Состояние системы в момент t определяется совокупностью состояний отдельных ее элементов в этот момент.

Если состояние i-го элемента в момент t обозначить через $s_i(t)$, то состояние системы

$$s(t)=(s_1(t), \ldots, s_n(t)).$$

Определенным совокупностям состояний элементов соответствует состояние исправности системы S (например, состояние $(s_1, s_2, ..., s_n)$ всегда есть состояние исправности для системы). Другим совокупностям состояний элементов соответствует состояние отказа системы $\bar{S}$ (например, состояние $(\bar{s}_1, \bar{s}_2, ..., \bar{s}_n)$ всегда есть состояние отказа системы).

Все множество состояний системы принято называть фазовым пространством состояний системы. В общем случае фазовое пространство не обязательно дискретно. Например, для описания элемента можно использовать текущее значение определяющего его непрерывного параметра. Более того, иногда элементы удается описывать только несколькими такими параметрами. Однако на практике, как правило, вводят поле допусков на параметры, т. е. критерий отказа: вся область возможных значений разбивается на две — допустимых (состояние работоспособности) и недопустимых (состояние отказа) значений.

Независимо от того, рассматриваем ли мы дискретное или непрерывное фазовое пространство при исследовании надежности, всегда остается одна существенная сторона этого фазового пространства: множество всех состояний четко подразделяется на состояния двух типов — работоспособности и отказа.

1.5. Показатели надежности

Поскольку поведение любого технического объекта во времени описывается случайным процессом переходов из состояний работоспособности в состояние отказа и обратно, показатели надежности носят вероятностно-временной характер.

Невосстанавливаемые объекты в первую очередь характеризуются наработкой до отказа. Как для всякой случайной величины, лем надежности вычислительной системы будет вероятность, что за отведенное на решение время t_0 найдется хотя бы один интервал безотказной работы, который превысит необходимую длительность работы всей программы θ_0. Во втором случае после устранения очередного отказа работа программы может быть продолжена, начиная с места остановки. При этом нас будет интересовать вероятность того, что суммарная наработка вычислительной системы в отведенном на решение времени t_0 окажется не меньше чем время, необходимое для завершения программы θ_0.

Большинство сложных вычислительных программ, осуществляющих длительное непрерывное решение задач, предусматривает возможность такой работы в рестартовом режиме. При этом реально оказывается, что продолжение света возможно не с момента возникновения отказа или сбоя в системе, а с момента последней регистрации контрольного состояния вычислительного процесса, т. е. работа, выполненная с момента последней регистрации до отказа, обесценивается. Безусловно, имеется возможность учесть эффект, связанный с частичным обесцениванием проделанной работы.

1.6. Выбор показателей надежности

Выбор показателей надежности определяется конкретной задачей, выполняемой системой, ее назначением и общими требованиями к результатам ее функционирования [5, 6]. Так, если система предназначена для работы в режиме дежурства, то наиболее естественно в качестве основного показателя надежности выбрать коэффициент оперативной готовности. Если выполняемые системой задачи кратковременны (практически мгновенны), то удобным показателем надежности оказывается коэффициент готовности. Если система перед началом работы допускает тщательный контроль работоспособности, но в процессе выполнения задачи не допускает восстановления, то удобным показателем надежности оказывается вероятность безотказной работы в течение времени выполнения задачи. При анализе эксплуатационных затрат важным показателем надежности оказывается средняя наработка на отказ.

Кроме того, во многих случаях показатели надежности удобно подразделять на внешние (оперативно-тактические, технические) и внутренние (в некотором смысле «технологические»). Первые используются для характеристики надежности объектов, выполняющих «внешние» задачи и характеризующихся определенным выходным эффектом. Вторые нужны в основном проектировщикам: для определения показателей надежности объектов, выполняющих независимые функции, необходимо знать некоторые специальные характеристики элементов. Поясним это на примере с дублированной системой.

Пользователю необходимо, чтобы вычислительный комплекс работал безотказно в течение выполнения относительно длительной задачи. Для повышения надежности предполагается использовать нагруженный резерв специального вида: задача параллельно выполняется на обеих ЭВМ, причем в качестве критерия правильности выполнения текущего решения используются программные и аппаратные средства, позволяющие на следующий этап решения передавать результат функционирования только правильно работающей ЭВМ. И итоге пользователю безразлично, каким путем достигнута требуемая надежность вычислительного комплекса, его интересуют значение коэффициента оперативной готовности каждой ЭВМ и более детальные характеристики (распределение случайной наработки до отказа, распределение длительности восстановления отказавшей ЭВМ), поскольку именно это позволяет построить необходимую математическую модель для вычисления требуемого внешнего показателя надежности вычислительного комплекса.

Интересно отметить, что выбор вида показателя надежности или совокупности таких показателей для системы зависит от типа системы (например, восстанавливаемая или невосстанавливаемая), от режима ее функционирования и вида выполняемых задач, но не зависит от того, насколько ответственны сами выполняемые задачи. Последнее определяющим образом сказывается на задаваемой норме надежности, но не на виде показателя.

При выборе показателей надежности следует иметь в виду достаточно очевидные соображения:

показатель надежности должен отражать способность системы выполнять свои основные функции;

нельзя задавать большое число показателей, поскольку это усложняет задачу оценки качества системы и пользователю, и самому разработчику;

выбранный показатель надежности должен иметь ясный физический смысл, причем следует избегать применения различных «сверток» показателей (типа средневзвешенных), имеющих к тому же различную физическую природу;

показатель надежности должен допускать возможность расчета его на этапе проектирования;

показатель надежности должен допускать возможность экспериментальной проверки его во время специальных испытаний или эксплуатации (возможно, и косвенным путем).

2. ВЕРОЯТНОСТНЫЕ МОДЕЛИ НАДЕЖНОСТИ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

2.1. Введение

На ранних этапах проектирования, когда бывают известны лишь общие принципы функционирования системы информатики и имеются лишь достаточно грубые представления о ее структуре, широко используются различные приближенные расчеты показателей надежности. Эти расчеты служат для того, чтобы дат оценку различным возможным вариантам построения создаваемо системы, т. е. на основе количественных оценок и сравнений выб рать наиболее в том или ином смысле рациональный вариан В этом случае обычно применяются достаточно простые и в боль шой степени идеализированные математические модели. Это оп равдывается тем, что обычно на ранних этапах проектировани лишь весьма приближенно известны необходимые для вероятност ных расчетов статистические данные о надежности различны элементов, а некоторые элементы (узлов, подсистем) создаютс параллельно с самой системой, поэтому о их надежности вообщ имеются лишь суждения экспертов.

Однако было бы неправильно считать, что вероятностные рас четы надежности нужны только на таких ранних стадиях проек тирования, когда осуществляется лишь почти качественное срав нение вариантов проектируемой системы. Позже, когда становитс известной статистическая информация о надежности самой созда ваемой системы по результатам различных стендовых испытани или опытных прогонов, вероятностные расчеты могут уже испол зоваться для получения объективных оценок показателей надеж ности, которые просто необходимы для разработки различных ор ганизационно-технических мероприятий по обеспечению надеж ности.

Ниже будут рассмотрены некоторые математические модел надежности, которые могут быть использованы при проектиров нии систем информатики.

2.2. Невосстанавливаемые системы

Системы информатики, в том числе информационно-вычисли тельные, предназначаются для функционирования в течение дл тельных периодов времени, а поэтому, конечно, в процессе экспл атации допускают проведение различного рода восстановительн мероприятий (ремонтов, регламентных профилактических раб и т. п.). Однако на определенных периодах эксплуатации п анализе надежности эти же системы рассматриваются как нево станавливаемые, например если в процессе выполнения какой-л бо задачи принципиально невозможно восстановление отказавш элементов и других частей рассматриваемой системы. Пожалу главной отличительной чертой невосстанавливаемой системы я ляется то, что при анализе ее надежности, как правило, предпол гается, что в начальный момент все элементы системы «соверше но новые» либо, как минимум, исправны.

Последовательная система. В теории надежности последов тельной называется такая система, отказ хотя бы одного элемен которой приводит к отказу системы в целом. Обычно рассматр ваются системы, состоящие из независимых элементов. В это случае для вероятности безотказной работы системы можно з писать

$$P(t) = \prod_{1\le i\le n} P_i(t),$$

где $P_i(t)$ — вероятность безотказной работы i-го элемента системы.

Средняя наработка до отказа системы, как правило, находится численными методами по формуле (1.1).

Если элементы системы имеют экспоненциальные распределения наработки до отказа с параметрами λ_i, то

$$P(t) = \exp\left(-\sum_{1\le i\le n} \lambda_i t\right),$$

$$T = \left(\sum_{1\le i\le n} \lambda_i\right)^{-1}.$$

Резервирование без восстановления. При проектировании систем информатики одним из эффективных и достаточно просто реализуемых методов повышения надежности является резервирование. Резервированием называется способ повышения надежности, заключающийся в использовании в составе системы дополнительных элементов, которые в случае отказа основных (рабочих) элементов могут брать на себя их функции.

В практике разработки средств вычислительной техники резервирование используется на самых различных структурных уровнях — от отдельных элементов устройств до процессоров и даже ЭВМ в целом. При этом резерв может находиться в таком же режиме, что и основные элементы, т. е. в такой же степени подвергаться отказам и расходовать свой ресурс, а может находиться в запасе. В первом случае говорят о нагруженном, а во втором — о ненагруженном резерве. Существует и промежуточный случай — облегченный резерв. Однако заметим, что на практике методы расчета систем с облегченным резервом развиты слабо из-за того, что они имеют относительно ограниченное прикладное значение: для расчетов надежности в этом случае не хватает ни статистической информации, ни реальных данных о степени разгруженности режима для такого резерва. Кроме того, удается построить более или менее приемлемую математическую модель облегченного резервирования только для экспоненциального распределения времени безотказной работы.

Рассмотрим сначала нагруженный резерв. Пусть основной элемент резервируется $n-1$ идентичными резервными элементами, каждый из которых находится в том же режиме, что и основной рабочий элемент. Случайная наработка системы до отказа

$$\xi = \max_{1\le i\le n} \xi_i,$$

где ξ_i — случайная наработка до отказа i-го элемента резервной группы (резервная группа включает в свой состав и основной элемент).

Если элементы взаимно независимы, то, используя теорему умножения вероятностей, можно записать выражение для функции распределения наработки системы до отказа (или, что то же самое, для вероятности отказа) в виде

$$F_\Sigma(t) = \mathscr{P}\{\xi \leqslant t\} = \prod_{1 \leqslant i \leqslant n} F_i(t).$$

Таким образом, для вероятности безотказной работы системы с нагруженным резервом

$$P_\Sigma(t) = 1 - \prod_{1 \leqslant i \leqslant n} [1 - P_i(t)],$$

где $P_i(t)$ — вероятность безотказной работы i-го элемента системы, $i = 0, 1, ..., n$.

Интуитивно понятно, что нагруженный резерв дает тем больший выигрыш, чем больше значение дисперсии распределения наработки до отказа. Действительно, если каждый элемент имеет вырожденное распределение наработки до отказа (т. е. наработка является постоянной величиной), то никакое увеличение кратности резервирования не может привести к выигрышу в надежности по сравнению с обычным одиночным элементом. В случае, когда все элементы одинаковы и имеют экспоненциальное распределение наработки до отказа, легко вычисляется средняя наработка до отказа всей системы:

$$T = \int_0^\infty [1 - (1 - e^{-\lambda t})^n]\, dt = \frac{1}{\lambda} \sum_{k=1}^{n+1} \frac{1}{k} .$$

На практике часто используется не поэлементное резервирование, а резервирование целыми подсистемами или системами. Пусть имеется последовательная система, состоящая из n элементов (для общности предположим, что все элементы разные). Рассмотрим два случая:

1) m-кратное общее резервирование, т. е. включаются параллельно m идентичных систем (рис. 2.1,a);

2) m-кратное поэлементное резервирование (рис. 2.1,$б$).

Для первого случая вероятности безотказной работы системы

$$P_\Sigma^{(1)}(t) = 1 - \left[1 - \prod_{i=1}^{n} P_i(t)\right]^m, \tag{2.1}$$

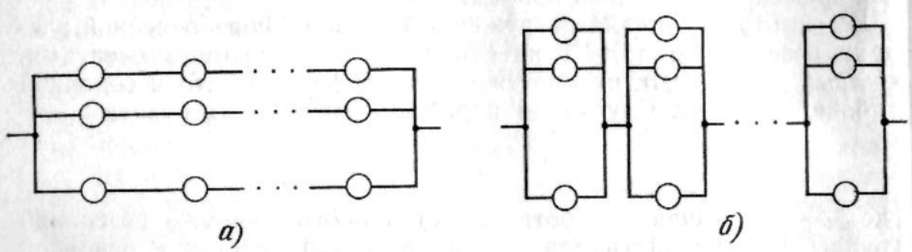

Рис. 2.1. Общее (a) и раздельное ($б$) резервирование

и для второго

$$P_\Sigma^{(2)}(t) = \prod_{i=1}^{n} \{1 - [1 - P_i(t)]^m\}. \tag{2.2}$$

Непосредственно из формул (2.1) и (2.2) видно, что поэлементное резервирование эффективнее общего. Для проверки отмеченного факта можно воспользоваться следующими очевидными рассуждениями.

Пусть ξ_{ij} — случайная наработка i-го элемента j-го участка системы. Тогда, учитывая, что наработка до отказа системы параллельно включенных элементов определяется максимальной из наработок элементов, а наработка системы последовательных элементов — минимальной из соответствующих величин, можно записать для случайной наработки до отказа рассматриваемых резервных систем:

$$\xi_\Sigma^{(1)} = \max_{1 \leqslant i \leqslant m} \min_{1 \leqslant j \leqslant n} \xi_{ij},$$

$$\xi_\Sigma^{(2)} = \min_{1 \leqslant j \leqslant n} \max_{1 \leqslant i \leqslant m} \xi_{ij},$$

откуда, поскольку $\min\max \geqslant \max\min$, сразу следует, что вероятность безотказной работы системы с общим резервированием меньше соответствующей вероятности системы с раздельным резервированием:

$$P_\Sigma^{(1)}(t) = \mathscr{P}\{\xi_\Sigma^{(1)} \geqslant t\} \leqslant P_\Sigma^{(2)}(t) = \mathscr{P}\{\xi_\Sigma^{(2)} \geqslant t\},$$

а также то, что для средних наработок выполняется условие

$$T_\Sigma^{(1)} = M\xi_\Sigma^{(1)} \leqslant T_\Sigma^{(2)} = M\xi_\Sigma^{(2)}.$$

Справедливо ради следует заметить, что на бумаге преимущества поэлементного (или, как его иногда называют, раздельного) резервирования по сравнению с общим выглядят гораздо привлекательнее, чем при практическом инженерном проектировании. Дело в том, что в формулах (2.1) и (2.2) совершенно не учитывается надежность реальных переключающих устройств, системы контроля работоспособности и других не менее важных факторов.

Однако даже идеальный поэлементный нагруженный резерв на самом деле характеризуется достаточно умеренной эффективностью. Поясним сказанное на следующем примере. Пусть с помощью нагруженного резерва мы намерены повысить среднюю наработку до отказа в десять раз. Оказывается, что для элементов, имеющих экспоненциальное распределение наработки до отказа, выполнение поставленной задачи потребует включения параллельно около 10 тыс. элементов! А для действительно существенного повышения надежности — в 100 раз — потребуется уже астрономическое число резервных элементов — почти 10^{40} штук!

Гораздо более эффективным во всех смыслах является ненагруженный резерв (конечно, если опять рассматривать идеализированную схему). Рассмотрим опять один основной элемент, к которому придаются n резервных, причем до отказа основного элемента все резервные находятся в ненагруженном режиме, т. е. они никак не расходуют свой ресурс, а при отказе основного элемента его функции мгновенно принимает на себя один из резервных, а остальные остаются в ненагруженном резерве.

Для такой резервной системы случайная наработка до отказа

$$\xi = \sum_{1 \le i \le n} \xi_i,$$

где ξ_i — случайная наработка i-го элемента до отказа. Заметим, что гипотеза о независимости наработок до отказа отдельных элементов в этом случе более оправдана.

Вероятность безотказной работы системы из идентичных элементов

$$P_\Sigma(t) = 1 - F^{*n}(t),$$

где F^{*n} — n-кратная свертка распределения $F(t)$, т. е.

$$F^{*n}(t) = \int_0^t F^{*(n-1)}(t-y)\,dF(y) = \int_0^t F(t-y)\,dF^{*(n-1)}(y).$$

(Выражение для системы из различных элементов записывается аналогично, но более громоздко.)

Средняя наработка системы до отказа находится в данном случае элементарно даже для разных и зависимых элементов:

$$T = M\left\{\sum_{i=0}^{n} \xi_i\right\} = \sum_{i=0}^{n} T_i.$$

Потенциальная эффективность ненагруженного резерва по сравнению с нагруженным очевидна даже при первом поверхностном анализе: увеличение средней наработки до отказа в десять раз требует увеличения числа резервных элементов в системе всего в десять раз, увеличение в 100 раз — соответственно увеличения числа резервных элементов также в 100 раз, т. е. средняя наработка до отказа растет линейно от числа резервных элементов.

Однако и в этом случае не стоит обольщаться: рассмотренная схема резервирования является идеализированной. Действительно, для того, чтобы подключить ненагруженный резервный элемент вместо нагруженного, нужны уже не только хорошее переключающее устройство и система непрерывного контроля работоспособности основного элемента (для своевременного определения момента подключения резерва): резервный элемент, находящийся в ненагруженном режиме, как правило, не может сразу войти в рабочий режим, для этого потребуется определенное время. В вычислительной системе для продолжения нормального функционирования необходимо сначала ввести резервный элемент в соответствующее «информационное» состояние. Поэтому зачастую, говоря о ненагруженном резерве, чаще всего имеют в виду обеспечение системы запасными элементами, т. е. предполагается не то, что

система будет непрерывно работать, а то, что она будет иметь возможность непрерывно снабжаться запасными элементами для проведения экстренных восстановительных мероприятий.

2.3. Учет надежности контрольно-переключающих устройств при резервировании без восстановления

При оценке надежности систем с резервированием существенную роль играет реальная надежность контролирующего работоспособность и переключающего устройства. Как правило, при расчетах надежности эти устройства не учитываются, что приводит к существенному завышению расчетных показателей по сравнению с теми, которые получаются на практике. Именно ненадежность переключающего устройства и ограничивает на практике потенциальные возможности повышения надежности.

Схему резервной группы из n элементов с контрольно-переключающим устройством (КПУ) можно представить в виде своеобразной системы с релейной обратной связью (рис. 2.2): при отсутствии сигнала нормального функционирования на выходе работающего в данный момент элемента с помощью КПУ подключается работоспособный резервный элемент.

Рассмотрим два основных типа КПУ: первый характеризуется вероятностью отказа q_k при k-м срабатывании (т. е. при подключении k-го резервного элемента), а второй — распределением наработки до отказа $q(t)$ в течение периода времени t. В обоих случаях будем считать, что КПУ одновременно анализирует и состояние работоспособности резервных элементов и что выбор исправного элемента при очередном подключении осуществляется автоматически.

Рассмотрим ненагруженный резерв без восстановления. Обозначим через $V_n(t)$ вероятность безотказной работы резервной группы из n элементов с КПУ первого типа.

По формуле полной вероятности можно записать

$$V_n(t) = P(t) + (1 - q_1)\int_0^t V_{n-1}(t-x)\,dF(x), \tag{2.3}$$

где $P(t)$ — вероятность безотказной работы основного элемента (элемента, стоящего на рабочей позиции) за время t; $F(t) = 1 - P(t)$.

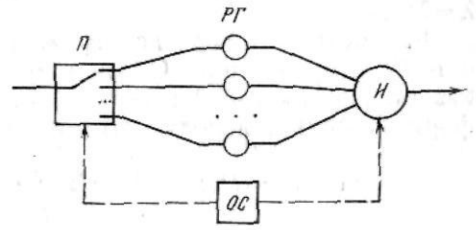

Рис. 2.2. Схема резервной группы элементов с контрольно-переключающим устройством: П — переключатель, РГ — резервная группа, И — индикатор работоспособности основного элемента, ОС — обратная связь (срабатывание переключателя на сигнал о неисправности основного элемента)

В этой рекуррентной формуле

$$V_2(t) = P(t) + (1 - q_{n-1}) \int_0^t V_1(t-x)\, dF(x),$$

$$V_1(t) = P(t).$$

Формула (2.3) означает, что для того, чтобы рассматриваемая резервная группа не отказала за время t, необходимо выполнение одного из двух событий:

1) не отказал за время t первый же основной элемент (т. е. резерв вовсе не понадобился);

2) основной элемент отказал в момент $0<x<t$, переключатель сработал нормально, а резервная группа теперь уже из $(n-1)$-го элемента благополучно проработает в течение оставшегося времени $t-x$.

Если положить $q_1 = \ldots = q_{n-1} = q$, а $p(t) = e^{-\lambda t}$, то последовательным интегрированием нетрудно найти

$$V_n(t) = e^{-\lambda t}\left(1 + (1-q)\lambda t + (1-q)^2 \frac{(\lambda t)^2}{2!} + \ldots + (1-q)^n \frac{(\lambda t)^n}{n!}\right) =$$

$$= e^{-\lambda t} \sum_{k=0}^{n} (1-q)^k \frac{(\lambda t)^k}{k!}.$$

Среднюю наработку до отказа можно также найти по рекуррентной формуле

$$T_n = T + (1-q)\, T_{n-1},$$

$$T_1 = T,$$

где

$$T = \int_0^\infty P(t)\, dt$$

— средняя наработка каждого элемента резервной группы.

Последовательной подстановкой получаем

$$T_n = T\,[1 + (1-q) + (1-q)^2 + \ldots + (1-q)^{n-1}] = \frac{T}{q}\,[1 - (1-q)^n]. \quad (2.4)$$

Для экспоненциального распределения формулу (2.4) можно получить особенно просто.

Заметим, что если при произвольных $P(t)$ n велико, а $q \ll 1$, то можно использовать экспоненциальное приближение для вероятности безотказной работы резервной группы, подставляя $\lambda = T^{-1}$.

Рассмотрим теперь ненагруженный резерв без восстановления с КПУ второго типа. Обозначим через $V_n(t)$ вероятность безотказной работы такой резервной группы из n элементов. Опять по формуле полной вероятности можно записать

$$\widetilde{V}_n(t) = P(t) + \int_0^t dq(x) \int_0^x P(t-y) \sum_{1 \le k \le n-1} dF^{*k}(y). \quad (2.5)$$

Эта формула означает, что для безотказной работы рассматриваемой резервной группы необходимо выполнение следующих событий:

1) за время t не отказал основной элемент;

2) в момент $0<x<t$ отказало КПУ, до этого момента отказало k элементов $(1 \le k \le n-1)$, причем последний из них отказал в момент $0<y<x$, а после этого последний включенный до отказа КПУ элемент безотказно проработал в течение времени $t-y$.

Если положить $P(t) = e^{-\lambda t}$, а $q(t) = 1 - e^{-\alpha t}$, то можно записать

$$\widetilde{V}_n(t) = e^{-\lambda t} + \int_0^t \alpha\, e^{-\alpha x} e^{-(t-x)\lambda} \sum_{k=1}^{n-1} \frac{(\lambda x)^k}{k!}\, dx =$$

$$= e^{-\lambda t}\left(1 + \alpha \sum_{k=1}^{n-1} \frac{\lambda^k}{k!}\left(\frac{k!}{(\alpha-\lambda)^{k+1}} - e^{-(\alpha-\lambda)t} \sum_{j=0}^{k!} \frac{k!}{j!} \frac{t^j}{(\alpha-\lambda)^{k-j+1}}\right)\right).$$

В этом случае среднюю наработку до отказа резервной группы можно найти приближенно из следующих соображений:

1) если отказ КПУ произошел до момента отказа последнего элемента резервной группы, то среднюю наработку до отказа можно оценить как

$$\widetilde{T}_n^{(1)} \approx T_{\text{КПУ}} + T',$$

где T' — остаточная средняя наработка элемента; $T_{\text{КПУ}}$ — средняя наработка КПУ;

2) если отказ всех элементов резервной групы произошел до отказа КПУ, то средняя наработка

$$\widetilde{T}_n^{(2)} \approx \min\left(\frac{n-1}{n} T_{\text{КПУ}},\ n\widetilde{T}'\right), \quad (2.6)$$

где $\widetilde{T}'$ — средняя наработка элемента при условии, что все элементы резервной группы отказали до отказа КПУ.

Вероятность второго из рассмотренных событий

$$\gamma^{(2)} = \int_0^\infty F^{*n}(t)\, dq(t),$$

а первого, естественно,

$$\gamma^{(1)} = 1 - \gamma^{(2)}.$$

Таким образом, средняя наработка резервной группы

$$\widetilde{T}_n \approx \gamma^{(1)} \widetilde{T}_n^{(1)} + \gamma^{(2)} \widetilde{T}_n^{(2)}.$$

Выше рассматривался случай, когда отказавшее КПУ не осуществляет переключения при отказе основного элемента. Однако в ряде схем работоспособность КПУ является обязательным условием работоспособности всей резервной группы. При этом учет надежности КПУ существенно упрощается. Отказ резервной группы наступает в том случае, если произойдет любое из двух событий:

1) отказ КПУ;

2) отказ всех элементов резервной группы.

Это фактически означает, что можно рассматривать последовательное соединение КПУ и соответствующей резервной группы, т. е.

$$V_n^*(t) = [1 - q(t)]\, V_n^0(t), \tag{2.7}$$

где V_n^* — вероятность безотказной работы резервной группы с КПУ; V_n^0 — вероятность безотказной работы резервной группы без учета КПУ.

Из формулы (2.7) следует, что надежность резервной группы ограничена надежностью КПУ. Более того, с ростом числа резервных элементов в группе переключатель естественным образом усложняется, что приводит в итоге к убыванию $V_n^*(t)$, начиная с некоторого n.

В этом случае средняя наработка находится обычным образом путем интегрирования (2.7).

2.4. Восстанавливаемая система (линейный граф)

На практике почти все технические системы после отказа восстанавливаются и продолжают свое функционирование. В математических моделях, используемых в инженерной практике, обычно предполагается, что подключение резервного элемента, если таковой в системе существует, происходит мгновенно, а отказавший элемент направляется для проведения восстановительных работ и по истечении случайного времени восстанавливает в результате ремонта работоспособность и поступает в систему опять в качестве резервного. Эти допущения, конечно, не всегда корректны при решении конкретных задач: каждый раз они должны обосновываться на содержательном уровне. Пример учета КПУ будет приведен ниже.

В общем случае для произвольной структуры системы, сложного процесса функционирования системы и произвольных законов распределения различных входящих в модель случайных величин построение математической модели процесса функционирования оказывается весьма сложным. Рассмотрим лишь марковскую модель, пояснив, какие допущения при этом необходимо сделать.

Предположим, известна структура системы и имеются вполне достоверные статистические данные о необходимых характеристиках функционирования (например, о распределениях наработки до отказа и времени восстановления отказавших блоков и элементов), о существующих возможностях проведения ремонта отказавших элементов и замены отказавших блоков и т. п. Тем не менее математическая модель рассматриваемого объекта, например многомашинного вычислительного комплекса, может оказаться настолько сложной, что аналитическое ее решение в строгой форме нельзя получить, а прямое статистическое моделирование на ЭВМ в случае высокой заданной надежности требует недопустимо большого времени. Более того, следует подчеркнуть, что проведение тонких и трудоемких математических расчетов на этапе предварительной оценки при весьма приближенных сведениях о параметрах проектируемой системы даже и не является оправданным.

В этом случае приходится прибегать к двум основным приемам: либо заменять точную математическую модель приближенной, в которой, например, произвольные функции распределения заменяются экспоненциальными (т. е. математическая модель объекта сводится к марковской, для которой хорошо известны методы решения, или же имеются готовые результаты в относительно компактной замкнутой форме), либо использовать асимптотические модели, в которых делаются предположения об определенных соотношениях между теми или иными параметрами системы.

Рассмотрим сначала расчет надежности идеализированной системы, состоящей из n однотипных элементов. Заметим, что несмотря на свою простоту, эта ситуация довольно часто встречается при исследовании вычислительных систем и систем передачи данных: в этих случаях ЭВМ, входящие в состав вычислительных систем, или линии связи в системах передачи данных однотипны или их можно в первом приближении считать таковыми. Число рассматриваемых однотипных элементов может быть достаточно большим. Из этих элементов n_1 являются основными, т. е. отказ каждого из них приводит к отказу системы, n_2 находятся в режиме нагруженного резерва, т. е. при прочих равных условиях они имеют показатели надежности, эквивалентные соответствующим характеристикам основных элементов, а n_3 элементов находятся в режиме облегченного резерва, т. е. они в меньшей степени подвержены отказам, чем основные элементы или резервные, находящиеся под нагрузкой.

Предполагается, что система построена так, что при отказе любого основного элемента на его место можно поставить любой элемент из нагруженного резерва, а место последнего занимает сразу же один из элементов, находящихся в облегченном резерве. Аналогично при отказе нагруженного резервного элемента на его место ставится элемент из облегченного резерва. Любой из отказавших элементов поступает на ремонт, который продолжается в течение случайного времени, зависящего от конкретных факторов (длительности непрерывной работы, после которой наступил отказ, режима работы, в котором находился элемент в момент отказа, и т. п.). Предположим, что одновременно может ремонтироваться не более двух отказавших элементов, т. е. при большем числе отказов наблюдается очередь на ремонт.

Заметим, что если даже известны функции распределения наработки до отказа системы в нагруженном режиме (т. е. для основных и нагруженных резервных элементов) $F(t)$ и функции распределения наработки до отказа элементов в облегченном режиме $F^*(t)$, то в общем случае этого недостаточно. Действительно, условное распределение наработки элемента, перешедшего из облегченного режима в нагруженный, может существенно зависеть от того, сколько времени до этого элемент уже проработал в облегченном режиме. Это связано с возможной выработкой ресурса и старением элементов в облегченном режиме. Как правило, подобной информации об изменении надежности элементов не имеется, поэтому построение какой-либо «точной» гипотетической модели может вообще не иметь смысла. Обычно нам неизвестно и то, в какой степени облегченный режим легче нагруженного, и даже то, относятся ли распределения $F(t)$ и $F^*(t)$ к одному классу.

В такой ситуации приходится делать некоторые упрощающие предположения в отношении законов распределения встречающихся в задаче случайных величин. Чаще всего в инженерной практике делаются предположения относительно экспоненциальности всех входящих в модель распределений случайных величин, т. е. рассматривают марковскую модель. Однако для того, чтобы использование марковской модели было более или менее обоснованным, следует проверить близость истинного распределения к экспоненциальному. В этом случае на практике не прибегают к слишком сложным критериям проверки: достаточно убедиться, что коэффициент вариации истинного распределения близок к единице, чтобы с большим основанием использовать для анализа функционирования рассматриваемой системы экспоненциальное распределение.

Если истинное распределение $F(t)$ считается известным по каким-то априорным сведениям или из каких-либо физических предпосылок, то математическое ожидание и дисперсию можно вычислить по стандартным формулам

$$T=\int_0^\infty x\,dF(x),\quad \sigma^2=\int_0^\infty (T-x)^2\,dF(x).$$

Если же известны результаты наблюдения за соответствующей случайной величиной $\theta_1, \theta_2, ..., \theta_N$, то среднее значение и выборочная дисперсия определяются по формулам

$$\hat{T}=\frac{1}{N}\sum_{i=1}^{N}\theta_i,\quad \hat{\sigma}^2=\frac{1}{N-1}\sum_{i=1}^{N}(\theta_i-\hat{T})^2.$$

Если при этом оказывается, что коэффициент вариации распределения, определяемый как

$$k=\sigma/T \text{ или } k=\hat{\sigma}/\hat{T},$$

сравнительно близок к единице, то можно считать, что гипотеза об экспоненциальности распределения $F(t)$ достаточно правдоподобна.

Далее для экспоненциального распределения характерно отсутствие последействия, т. е., сделав допущение о приемлемости этого распределения, мы тем самым снимаем вопрос о виде условного распределения наработки до отказа для элемента, например для элемента, проработавшего некоторое время в облегченном режиме перед переходом в нагруженный режим.

Осталось лишь заметить, что, по нашим предположениям, интенсивность отказов для облегченного режима меньше, чем для нагруженного. Чтобы зафиксировать этот факт, обозначим $\lambda^*=\nu\lambda$, где ν — коэффициент нагрузки, принимающий значения $0<\nu<1$.

Теперь нетрудно строго сформулировать марковскую модель для рассматриваемой восстанавливаемой системы, решение которой не представляет трудностей, а полученные результаты, как можно ожидать, будут достаточно приемлемыми для ориентировочных оценок.

Рассмотрим произвольное состояние системы, например (N, k), где $N=n_1+n_2$ и $0\leqslant k\leqslant n_3$, и запишем для него следующее уравнение, описывающее полную группу событий:

$$\begin{aligned} P(N, k; t+\Delta t) &= P(N, k; t)\,p(N, k|N, k; \Delta t)+\\ &+P(N, k+1; t)\,p(N, k|N, k+1; \Delta t)+\\ &+P(N, k-1; t)\,p(N, k|N, k-1; \Delta t),\end{aligned} \tag{2.8}$$

где $P(X, Y; z)$ — вероятность того, что система находится в состоянии (X, Y) в момент времени z; $p(X, YX^*, Y^*; \Delta z)$ — условная вероятность перехода из состояния (X^*, Y^*) в состояние (X, Y) за время Δz.

Нетрудно записать выражение для условных вероятностей перехода:

$$\begin{aligned} p(N, k|N, k; \Delta t) &= 1-[2\mu+(N+k\nu)\lambda]\Delta t,\\ p(N, k|N, k+1; \Delta t) &= [N+(k+1)\nu]\lambda\Delta t,\\ p(N, k|N, k-1; \Delta t) &= [N+(k-1)\nu]\lambda\Delta t,\end{aligned} \tag{2.9}$$

где μ — интенсивность восстановления.

Подставив (2.9) в (2.8), после элементарных преобразований и предельного перехода при $\Delta t\to 0$ получаем дифференциальное уравнение вида

$$\begin{aligned} P'_{N,k}(t) &= \lambda[N+(k+1)\nu]P_{N,k+1}(t)-\\ &-[2\mu+\lambda(N+k\nu)]P_{N,k}(t)+\lambda[N+(k-1)\nu]P_{N,k-1}(t).\end{aligned}$$

Полная система дифференциальных уравнений, описывающая функционирование рассматриваемой системы, имеет вид

$$P'_{N,n_3}(t)=-\lambda(N+\nu n_3)P_{N,n_3}(t)+\mu P_{N,n_3-1}(t),$$

$$\begin{aligned} P'_{N,k}(t) &= \lambda[N+(k+1)\nu]P_{N,k+1}(t)-(2\mu+\lambda)(N+k\nu)]P_{N,k}(t)+\\ &+\lambda[N+(k-1)\nu]P_{N,k-1}(t) \quad \text{для} \quad k>0,\end{aligned}$$

$$P'_{m,0}(t) = \lambda(m+1)P_{m+1,0}(t) - (2\mu + \lambda m)P_{m,0}(t) + \lambda(m-1)P_{m-1,0}(t)$$

$$\text{для} \quad 0 < m < N,$$

. . . .

$$P'_{0,0}(t) = \lambda P_{1,0}(t) - 2\mu P_{0,0}(t).$$

В качестве начальных условий можно взять, например, $P_{N,n_3}(0) = 1$, т. е. при $t=0$ в системе нет ни одного отказавшего элемента (возможны и другие начальные условия). Эта система уравнений известна под названием схемы гибели и размножения или рождения и смерти. (Название объясняется тем, что подобные уравнения впервые появились в математических моделях в демографии.)

Такая схема удобно описывается при помощи линейного графа переходов (рис. 2.3).

Нас могут интересовать самые разные показатели надежности рассматриваемой системы: вероятность безотказной работы $P(t)$ в интервале времени $(0, t)$, средняя наработка до первого отказа или между отказами в стационарном (установившемся) режиме, коэффициент готовности в стационарном режиме и др. Остановимся на нахождении одного из простейших показателей — коэффициента готовности, т. е. вероятности того, что в произвольный «достаточно удаленный» момент времени система будет находиться в работоспособном состоянии, т. е. в ремонте одновременно будет находиться не более n_2+n_3 элементов.

При вычислении коэффициента готовности рассматривается стационарный режим, а это означает, что система дифференциальных уравнений переходит в систему линейных алгебраических уравнений (стоящие в левых частях уравнений производные обращаются в нуль, а все $P_{ij}(t)$ обращаются в константы с теми же индексами P_{ij}).

Решение полученной линейной системы алгебраических уравнений, которая соответствует схеме гибели и размножения, хорошо известной в теории массового обслуживания и теории надежности.

Введем обозначения

$$\theta_{N,k} = \frac{\lambda^{n_3-k} \prod_{i=0}^{k-1} [N + \nu(n_3 - i)]}{\mu^{n_3-k} 2^{n_3-k-1}} \quad \text{при} \quad k > 0,$$

$$\theta_{m,0} = \theta_{N,0} = \frac{(\nu\lambda)^{N-m} \prod_{i=0}^{N-m-1} (m+i)}{\mu^{N-m} 2^{N-m}} \quad \text{при} \quad m \leqslant N$$

и далее

$$\theta = \sum_{k=1}^{n_3} \theta_{N,k} + \sum_{m=0}^{N} \theta_{m,0}.$$

Тогда для вероятностей состояний можно записать

$$P_{N,k} = \frac{\theta_{N,k}}{\theta}, \quad P_{m,0} = \frac{\theta_{m,0}}{\theta}.$$

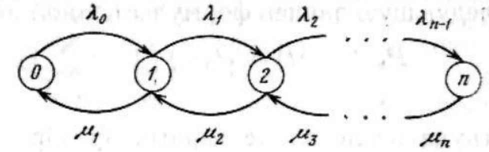

Рис. 2.3. Линейный граф переходов (схема «гибели и размножения»)

Работоспособными состояниями системы являются (N, n_3), (N, n_3-1), ..., $(N,0)$... $(n_1+1,0)$, $(n_1, 0)$. Следовательно, коэффициент готовности равняется сумме вероятностей пребывания во всех этих состояниях:

$$K = \sum_{k=1}^{n_3} P_{N,k} + \sum_{m=n_1}^{N} P_{m,0}.$$

Заметим, что ряд стационарных показателей надежности типа коэффициента готовности, вычисленных в предположении об экспоненциальности распределений, сохраняет тот же вид и в более общем случае.

2.5. Восстанавливаемая система (граф общего вида)

Большинство реальных систем описываются сложными графами переходов, для них возможны сложные фазовые пространства состояний, сложные критерии принадлежности их состояний к работоспособности или отказу.

Однако для марковских моделей и в этом случае принципиальных трудностей при составлении математического описания и последующего решения (как правило, численного) не возникает. Покажем в общем виде, как это осуществляется [5].

Если известны или заданы описание алгоритма функционирования и правила восстановления элементов системы, то можно определить, в каких состояниях она может находиться. Задав определенный критерий отказа, все состояния системы можно подразделить на два класса: работоспособности и отказа. По известному описанию системы и известным количественным показателям надежности отдельных элементов системы и длительности их ремонта может быть построен граф переходов, у которого вершинами будут состояния системы, а дугами — переходы с интенсивностями, определяемыми соответствующими интенсивностями отказов и восстановлений элементов.

Пусть граф переходов, описывающий процесс функционирования системы, построен. Если этот граф имеет n различных состояний, то для получения различных показателей надежности в общем случае потребуется выписать систему из n уравнений. Рассмотрим некоторое состояние k, в которое можно попасть из некоторого множества состояний G_1^k и из которого, в свою очередь, можно попасть в одно из состояний множества G_2^k. Дифференцированное уравнение для данного состояния получим, используя следующую запись формулы полной вероятности:

$$P_k(t+\Delta t)=P_k(t)\left(1-\sum_{i\in G_2^k}\lambda_{ki}\Delta t\right)+\sum_{i\in G_1^k}P_i(t)\lambda_{ik}\Delta t,$$

откуда после элементарных преобразований и перехода к пределу при $\Delta t\to 0$ получаем

$$P_k'(t)=-P_k(t)\sum_{i\in G_2^k}\lambda_{ki}+\sum_{i\in G_1^k}\lambda_{ik}P_i(t).$$

Для системы с общим числом состояний N требуется составить систему таких линейных дифференциальных уравнений, в которую следует, исключив одно из уравнений, ввести условие нормировки (уравнение полной группы событий)

$$\sum_{1\leq i\leq N}P_i(t)=1.$$

При этом, если определяются показатели типа нестационарного коэффициента готовности, то строится граф переходов со всеми возможными переходами из одного состояния в другое. Если же отыскивается вероятность безотказной работы в течение некоторого интервала времени или же средняя наработка до отказа, то необходимо все состояния отказа сделать поглощающими, т. е. обратить соответствующие интенсивности переходов из этих состояний в ноль (обратить интенсивности восстановления из состояний отказа системы в ноль).

Понятно, что в этом случае последняя система уравнений примет вид

$$P_k'(t)=-P_k(t)\sum_{i\in G_2^k}\lambda_{ki}+\sum_{i\in \widetilde{G}_1^k}\lambda_{ik}P_i(t),\quad k=1,N,$$

$$\sum_{1<i<N}P_i(t)=1,$$

где множество $\widetilde{G}_1^k$ является подмножеством множества G_1^k, из которого исключены все поглощающие состояния (состояния отказа).

Записанные системы уравнений можно решить любым удобным способом. В последние годы в инженерной практике чаще всего используется метод преобразований Лапласа—Стилтьеса. Этот метод удобен сам по себе и позволяет получать выражения, которые дают возможность исследовать различные асимптотические (предельные) свойства рассматриваемых систем.

Для нахождения характеристик на основании системы дифференциальных уравнений сначала следует применить к ним преобразования Лапласа и найти решение в преобразованиях. Далее, просуммировав все преобразования Лапласа по интересующему нас классу состояний (например, по состояниям работоспособности), находим преобразование Лапласа для искомой вероятности. Разложив полученное выражение на простые дроби, найдем решение. Если получено преобразование Лапласа для вероятности безотказной работы, то для нахождения средней наработки до отказа достаточно найти значение данного преобразования при $s=0$.

Для определения стационарных характеристик в этом же случае получаем следующее алгебраическое уравнение:

$$0=-P_k\sum_{i\in G_2}\lambda_i+\sum_{i\in G_1}\lambda_iP_i.$$

Требуемая стационарная вероятность находится как сумма вероятностей состояний, принадлежащих к интересующему нас подмножеству состояний (по состояниям работоспособности, если отыскивается коэффициент готовности, или по состояниям отказа, если отыскивается коэффициент простоя).

2.6. Дублированная система (экспоненциальное распределение времени восстановления)

Самой простой резервированной системой является дублированная система, в которой имеется один основной элемент и один резервный. Простота этой системы позволяет получить многие интересные результаты в удобной форме, а также показать ряд полезных математических приемов, которые приводят к хорошим результатам и для более сложных систем, но здесь эти приемы могут быть продемонстрированы более понятно и, может быть, даже более эффективно.

Для дублированной системы из идентичных элементов решения в замкнутой форме могут быть получены в общем виде. Граф переходов для дублированной системы представлен на рис. 2.4. Понятно, что различные режимы резерва и восстановления элементов приводят лишь к различным соотношениям между интенсивностями отказов и восстановлений. Поясним сказанное для четырех основных схем дублирования с восстановлением (в предположении экспоненциальности всех распределений времени пребывания в отдельных состояниях).

1. Нагруженный резерв, неограниченное восстановление. В этом случае предполагается, что резервный элемент находится в том же режиме, что и основной, т. е. интенсивность его отказов равна интенсивности отказов основного элемента, причем при отказе обоих элементов возможно одновременное и независимое их восстановление. Таким образом, для этой модели

$$\lambda_0=2\lambda,\ \lambda_1=\lambda,\ \mu_1=\mu,\ \mu_2=2\mu.$$

2. Нагруженный резерв, ограниченное восстановление. Этот случай отличается от предыдущего режимом восстановления: при отказе двух элементов восстанавливается лишь один из них. Заметим, что в силу предположения об идентичности элементов и

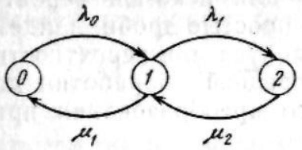

Рис. 2.4. Граф переходов для дублированной системы

экспоненциальности распределения наработки до отказа не играет роли дисциплина обслуживания: остаточное время дообслуживания того элемента, который уже восстанавливался к моменту отказа второго элемента, ничем в вероятностном смысле не отличается от времени обслуживания только что отказавшего элемента. Для этой модели следует положить

$$\lambda_0=2\lambda,\ \lambda_1=\lambda,\ \mu_1=\mu_2=\mu.$$

3. Ненагруженный резерв, неограниченное восстановление. Эта схема отличается от первой режимом работы резервного элемента: до момента отказа резервный элемент отказать не может, т. е. его интенсивность отказов при нахождении в резерве равна нулю. В этом случае

$$\lambda_0=\lambda_1=\lambda,\ \mu_1=\mu,\ \mu_2=2\mu.$$

4. Ненагруженный резерв, ограниченное восстановление. Учитывая предыдущие комментарии, можно сразу без пояснений написать

$$\lambda_0=\lambda_1=\lambda,\ \mu_1=\mu_2=\mu.$$

После этих замечаний рассмотрим получение различных показателей надежности для дублированной системы. При определении характеристик типа вероятности безотказной работы, средней наработки до отказа, коэффициента оперативной готовности состояние «2» (рис. 2.4) следует считать поглощающим, т. е. полагать $\mu_2=0$.

Используя общую схему составления алгебраических уравнений, представляющих собой уравнения равновесия, можем для стационарного коэффициента готовности написать

$$-\lambda_0P_0+\mu_1P_1=0,$$
$$\lambda_0P_0-(\lambda_1+\mu_1)P_1+\mu_2P_2=0,$$
$$P_0+P_1+P_2=1.$$

Поскольку для дублированной системы состояниями работоспособности являются состояния «0» и «1», коэффициент готовности для этой системы равен P_0+P_1. Решение записанной выше системы уравнений легко находится по правилу Крамера:

$$K=1-P_3=1-\begin{vmatrix}-\lambda_0 & \mu_1 & 0\\ \lambda_0 & -(\lambda_1+\mu_1) & 0\\ 1 & 1 & 1\end{vmatrix}\Bigg|\begin{vmatrix}-\lambda_0 & \mu_1 & 0\\ \lambda_0 & -(\lambda_1+\mu_1) & \mu_2\\ 1 & 1 & 1\end{vmatrix}=$$
$$=\frac{\lambda_0\lambda_1}{\lambda_0\lambda_1+\lambda_0\mu_2+\mu_1\mu_2}.$$

Для нахождения нестационарного коэффициента готовности, представляющего собой вероятность того, что дублированная система будет находиться в состоянии работоспособности в момент t при условии, что в момент $t=0$ она находилась в состоянии работоспособности, по графу переходов нетрудно непосредственно составить следующую систему линейных дифференциальных уравнений:

$$P_0'(t)=-\lambda_0P_0(t)+\mu_1P_1(t),$$
$$P_1'(t)=\lambda_0P_0(t)-(\lambda_1+\mu_1)P_1(t)+\mu_2P_2(t),$$
$$1=P_0(t)+P_1(t)+P_2(t),$$
$$P_0(0)=1.$$

Преобразование Лапласа для этой системы записывается как

$$(\lambda_0+s)\varphi_0(s)-\mu_1\varphi_1(s)=1,$$
$$-\lambda_0\varphi_0(s)+(s+\lambda_1+\mu_1)\varphi_1(s)-\mu_2\varphi_2(s)=0,$$
$$s\varphi_0(s)+s\varphi_1(s)+s\varphi_2(s)=1.$$

Решение имеет вид

$$\varphi_2(s)=\frac{\begin{vmatrix}s+\lambda_0 & -\mu_1 & 1\\ -\lambda_0 & s+\lambda_1+\mu_1 & 0\\ s & s & 1\end{vmatrix}}{\begin{vmatrix}s+\lambda_0 & -\mu_1 & 0\\ -\lambda_0 & s+\lambda_1+\mu_1 & -\mu_2\\ s & s & s\end{vmatrix}}=\lambda_1\lambda_0\{s[s^2+(\lambda_0+\lambda_1+\mu_1+\mu_2)s+$$
$$+\lambda_0\lambda_1+\lambda_0\mu_2+\mu_1\mu_2]\}^{-1}.$$

Найдем корни знаменателя:

$$s_{1,2}=-\frac{\lambda_0+\lambda_1+\mu_1+\mu_2}{2}\pm$$
$$\pm\sqrt{\frac{(\lambda_0+\lambda_1+\mu_1+\mu_2)^2}{4}-\lambda_0\lambda_1-\lambda_0\mu_2-\mu_1\mu_2}.$$

Таким образом,

$$\varphi_2(s)=\frac{\lambda_0\lambda_1}{s(s-s_1)(s-s_2)}.$$

Разложим выражение для $\varphi_2(s)$ на простые дроби:

$$\varphi_2(s)=\frac{A}{s}+\frac{B}{s-s_1}+\frac{C}{s-s_2}.$$

Если мы определим A, B и C, то $\varphi_2(s)$ легко обратить по стандартным формулам обратного преобразования Лапласа. Предварительно приведем последнее выражение к виду

$$\varphi_2(s)=\frac{A(s-s_1)(s-s_2)+Bs(s-s_2)+Cs(s-s_1)}{s(s-s_1)(s-s_2)}=$$

$$=\frac{s^2(A+B+C)-[A(s_1+s_2)+Bs_2+Cs_1]\,s+As_1s_2}{s(s-s_1)(s-s_2)}.$$

Для нахождения этих коэффициентов воспользуемся тем фактом, что равенство числителей предполагает равенство коэффициентов при соответствующих степенях s обоих полиномов. Это позволит записать следующую систему уравнений:

$$A=\lambda_0\lambda_1/(s_1s_2).$$
$$A(s_1+s_2)+Bs_2+Cs_1=0,$$
$$As_1s_2=\lambda_0\lambda_1.$$

Отсюда

$$A=\lambda_0\lambda_1/(s_1s_2),$$

и система уравнений примет вид

$$B+C=-\frac{\lambda_0\lambda_1}{s_1s_2},\ Bs_2+Cs_1=-\frac{\lambda_0\lambda_1}{s_1s_2}(s_1+s_2).$$

Из нее найдем

$$B=\frac{\lambda_0\lambda_1}{s_1(s_1-s_2)},\quad C=\frac{\lambda_0\lambda_1}{s_2(s_2-s_1)}.$$

Таким образом, обращение дает

$$P_2(t)=\frac{\lambda_0\lambda_1}{s_1s_2}+\frac{\lambda_0\lambda_1}{s_1(s_1-s_2)}\mathrm{e}^{-s_1t}+\frac{\lambda_0\lambda_1}{s_2(s_2-s_1)}\mathrm{e}^{-s_2t}$$

или

$$P_2(t)=\frac{\lambda_0\lambda_1}{s_1s_2}\left[1-\frac{1}{s_1-s_2}(s_1\mathrm{e}^{-s_2t}-s_2\mathrm{e}^{-s_1t})\right].$$

Заметим, что нас интересовало нахождение нестационарного коэффициента готовности, а мы, по существу, нашли нестационарный коэффициент простоя, т. е. вероятность нахождения системы в состоянии отказа. Поэтому

$$K(t)=1-P_2(t),$$

где $P_2(t)$ — уже найденное выражение.

Вероятность безотказной работы дублированной системы с восстановлением находится из той же системы линейных дифференциальных уравнений, но с учетом того, что состояние «2» является в данном случае поглощающим, т. е. что $\mu_2=0$. При этом система линейных дифференциальных уравнений имеет вид

$$P_0'(t)=-\lambda_0P_0(t)+\mu_1P_1(t),$$
$$P_1'(t)=\lambda_0P_0(t)-(\lambda_1+\mu_1)P_1(t),$$
$$P_0(0)=1.$$

Ее преобразование Лапласа

$$(\lambda_0+s)\varphi_0(s)-\mu_1\varphi_1(s)=1,$$
$$-\lambda_0\varphi_0(s)+(s+\lambda_1+\mu_1)\varphi_1(s)=0,$$

откуда

$$\varphi_0(s)+\varphi_1(s)=\frac{\begin{vmatrix}1 & -\mu_1\\ 0 & \lambda_1+\mu_1+s\end{vmatrix}+\begin{vmatrix}\lambda_0+s & 1\\ -\lambda_0 & 0\end{vmatrix}}{\begin{vmatrix}\lambda_0+s & -\mu_1\\ -\lambda_0 & \lambda_1+\mu_1+s\end{vmatrix}}=$$

$$=(s+\lambda_0+\lambda_1+\mu_1)\,(s^2+s(\lambda_0+\lambda_1+\mu_1)+\lambda_0\lambda_1)^{-1}.$$

Как и раньше, находим корни знаменателя:

$$s_{1,2}=-\alpha/2\pm\sqrt{\alpha^2/4-\beta},$$

где $\alpha=\lambda_0+\lambda_1+\mu_1$; $\beta=\lambda_0\lambda_1$.

Из условия

$$\frac{s+\lambda_0+\lambda_1+\mu_1}{s^2+s(\lambda_0+\lambda_1+\mu_1)+\lambda_0\lambda_1}=$$

$$=\frac{A(s-s_2)+B(s-s_1)}{(s-s_1)(s-s_2)}=\frac{(A+B)s-As_2-Bs_1}{(s-s_1)(s-s_2)}$$

составляем систему уравнений

$$A+B=1,$$
$$-As_1-Bs_2=\alpha,$$

откуда

$$A=(s_2+\alpha)/(s_2-s_1);\ B=(s_1+\alpha)/(s_1-s_2).$$

Учитывая, что $s_1+s_2=-\alpha$, получаем окончательно

$$A=s_1/(s_1-s_2),\ B=-s_2/(s_1-s_2).$$

Итак,

$$P(t)=\frac{1}{s_1-s_2}(s_1\mathrm{e}^{-s_2t}-s_2\mathrm{e}^{-s_1t}).$$

Это позволяет непосредственно находить выражение для средней наработки системы до отказа на основании выражения для преобразования Лапласа при $s=0$:

$$T=\left.\frac{s+\lambda_0+\lambda_1+\mu_1}{s^2+(\lambda_0+\lambda_1+\mu_1)s+\lambda_0\lambda_1}\right|_{s=0}=$$

$$=\frac{\lambda_0+\lambda_1+\mu_1}{\lambda_0\lambda_1}=\frac{1}{\lambda_1}+\frac{1}{\lambda_0}+\frac{\mu_1}{\lambda_0\lambda_1}.$$

Подставив соответствующие значения λ_0, λ_1, μ_1 и μ_2, можно получить показатели надежности для всех описанных выше частных случаев. При этом для некоторых схем (а именно когда $s_1=s_2$) приходится определять показатели надежности, применяя правило Лопиталя.

2.7. Дублированная система (произвольное распределение времени восстановления)

Пусть время восстановления имеет произвольное распределение $G(x)$. Вероятность безотказной работы основного и резервного элементов подчиняется экспоненциальному закону с параметрами λ и λ_1 соответственно ($\lambda_1=\alpha\lambda$). Коэффициент α ($0\leqslant\alpha\leqslant1$) характеризует степень нагруженности режима резервного элемента. Пусть при отказе резервный элемент мгновенно становится на место основного [7[1]].

Безотказность системы в течение времени t определяется выполнением хотя бы одного из следующих событий:

ни один из элементов не откажет за время t;

один из элементов отказал в момент x ($0\leqslant x\leqslant t$), восстанавливался в течение времени y ($y>t-x$), второй элемент (всегда основной) в течение оставшегося времени $t-x$ не отказал;

один из элементов отказал в некоторый момент x ($0\leqslant x\leqslant t$), восстанавливался в течение времени y ($y<t-x$), в течение этого времени основной элемент не отказал, с момента $x+y$ до момента t, т. е. в течение оставшегося времени $t-x-y$, дублированная система проработает безотказно.

Вероятность безотказной работы системы за время t можно определить по формуле полной вероятности:

$$P(t)=e^{-\lambda(1+\alpha)t}+\int_0^t\lambda(1+\alpha)e^{-\lambda(1+\alpha)x}e^{-\lambda(t-x)}\times$$

$$\times[1-G(t-x)]dx+\int_0^t P(t-x)dx\int_0^x\lambda(1+\alpha)e^{-\lambda(1+\alpha)y-\lambda(x-y)}dG(x-y)=$$

$$=e^{-\lambda(1+\alpha)t}+\int_0^t\lambda(1+\alpha)e^{-\lambda(t+\alpha)x}[1-G(t-x)]dx+$$

$$+\int_0^t P(t-x)dx\int_0^x\lambda(1+\alpha)e^{-\lambda(x+\alpha y)}dG(x-y). \qquad (2.10)$$

Обозначим преобразования Лапласа

$$\varphi(s)=\int_0^\infty P(t)e^{-st}dt,\quad g(s)=\int_0^\infty e^{-st}dG(t).$$

После несложных преобразований получим решение уравнений (2.10) в терминах преобразований Лапласа в виде

$$\varphi(s)=\frac{\dfrac{s+\lambda+\lambda(1+\alpha)[1-g(s+\lambda)]}{[s+\lambda(1+\alpha)](s+\lambda)}}{1-\dfrac{\lambda(1+\alpha)g(s+\lambda)}{s+\lambda(1+\alpha)}}$$

[1] См. также: **Райкин А. Л.** Вероятностные модели функционирования резервированных устройств. — М.: Наука, 1971.

или после упрощений

$$\varphi(s)=\frac{s+\lambda+\lambda(1+\alpha)[1-g(s+\lambda)]}{(s+\lambda)\{s+\lambda(1+\alpha)[1-g(s+\lambda)]\}}.$$

Обращение полученного преобразования Лапласа в общем случае не приводит к выражению для вероятности безотказной работы $P(t)$ в замкнутой форме. Однако из $\varphi(s)$ можно найти выражение для средней наработки до отказа рассматриваемой системы, учтя, что

$$\varphi(0)=\int_0^\infty P(t)\,dt=T.$$

Окончательно получаем

$$T=\frac{\lambda+\lambda(1+\alpha)[1-g(\lambda)]}{\lambda^2(1+\alpha)[1-g(\lambda)]}=\frac{1}{\lambda}+\frac{1}{\lambda(1+\alpha)[1-g(\lambda)]}.$$

Заметим, что выражение

$$1-g(\lambda)=1-\int_0^\infty e^{-\lambda t}dG(t)=\int_0^\infty(1-e^{-\lambda t})\,dG(t)$$

имеет простой физический смысл — это вероятность того, что основной элемент не откажет за время восстановления.

Если среднее время восстановления τ много меньше средней наработки до отказа основного элемента T, т. е.

$$\frac{1}{\lambda}\gg\int_0^\infty t\,dG(t)=\tau,$$

то можно, использовав разложение для экспоненты, записать

$$1-g(\lambda)=\int_0^\infty\left(\lambda t-\frac{(\lambda t)^2}{2!}+\ldots\right)dG(t)\approx\lambda\tau-0{,}5\lambda^2(\tau^2+\sigma_\tau^2),$$

где σ_τ^2 — дисперсия распределения времени восстановления.

Если $g(\lambda)\ll1$, то можно найти асимптотическое распределение $P(t)$, используя теорему Реньи о разрежении потока восстановления. Потоком восстановления называется поток событий, случайные интервалы между которыми взаимно независимы и одинаково распределены. Теорема Реньи [5] гласит, что если к любому потоку восстановления применять операцию разрежения, которая является марковским исключением точек потока (с некоторой постоянной вероятностью, не зависящей от предыстории, исключаются точки исходного потока), то образуемый в результате регулярного применения указанной процедуры поток асимптотически стремится к пуассоновскому (при соответствующей нормировке). Теоремой Реньи можно воспользоваться для получения приближенных выражений, если вероятность исключения точек близка к единице.

Для дублированной системы интервал θ состоит из двух частей: *случайной наработки дублированной системы* от момента восстановления до отказа любого из элементов и случайного времени восстановления отказавшего элемента (при условии, что длительность восстановления меньше наработки основного элемента). По окончании такого цикла «работа—ремонт» дублированная система переходит в начальное состояние, когда оба элемента исправны. *Последний цикл, на котором происходит отказ дублированной* системы, отличается от предыдущих тем, что время восстановления оказавшего элемента оказывается больше наработки до отказа основного элемента. В этом случае наступает отказ дублированной системы в момент отказа основного элемента. (При этом следует учитывать, что распределение наработки основного элемента *вычисляется при упомянутом условии.*)

В реальных ситуациях когда рассматриваются высоконадежные дублированные системы, этими рассуждениями можно пренебречь: обычно вторая часть интервала θ — среднее время восстановления отказавшего элемента — много меньше средней наработки системы до появления любого отказа (а условная величина, ограниченная дополнительно сверху, будет *и подавно* меньше). Это позволяет считать, что

$$M\theta\approx[\lambda(1+\alpha)]^{-1}.$$

Таким образом, для высоконадежной дублированной системы с восстановлением при экспоненциальных распределениях наработок на отказ и произвольном распределении времени восстановления можно рассматривать по существу пуассоновский поток практически «точечных» событий. В этом случае разрежение потока приводит к пуассоновскому потоку всегда, т. е. даже в допредельном случае. Иначе говоря, условие $\lambda\tau\ll1$ обеспечивает автоматически приемлемость аппроксимации распределения наработки системы экспоненциальным распределением.

Если среднее расстояние между точками исходного потока равно θ, а операция разрежения осуществлялась с вероятностью r, то в полученном потоке после разрежения среднее расстояние между оставшимися точками

$$\theta^*=\theta/(1-r).$$

Для дублированной системы разрежение означает, что во время восстановления отказавшего элемента не произойдет отказа работающего в это время основного элемента. Вероятность этого события

$$r=\int_0^\infty e^{-\lambda t}dG(t).$$

Нетрудно видеть, что она численно равна $g(\lambda)$.

В то же время с учетом облегченного режима резервного элемента интенсивность возникновения отказов любого элемента дублированной системы $\lambda+\lambda_1=\lambda(1+\alpha)$.

Таким образом, вероятность безотказной работы дублированной системы *в данном случае может быть приближенно* записана в виде

$$P(t)\approx\exp\{-\lambda(1+\alpha)[1-g(\lambda)]\}.$$

Эта формула дает хорошее приближение, когда выполняется условие $g(\lambda)\ll1$. Обычно это условие эквивалентно условию $T\gg\tau$.

2.8. Учет надежности контрольно-переключающих устройств при резервировании с восстановлением

Рассматривая резерв с восстановлением, имеет смысл и КПУ считать восстанавливаемым. Возможны три основные схемы восстановления КПУ:

1) восстановление при выявлении невозможности переключения;

2) периодическая проверка КПУ через определенное время и восстановление только при обнаружении отказа в процессе такой проверки;

3) восстановление сразу после возникновения отказа.

Мы не будем рассматривать точных математических моделей для всех этих случаев, так как они очень громоздки. Рассмотрим только схему дублирования при условии высокой надежности и самих элементов резервной группы, и КПУ.

Опять, как и для резервирования без восстановления, рассмотрим КПУ двух типов: первый характеризуется вероятностью отказа в момент подключения резервного элемента, а второй — вероятностью безотказной работы во времени.

Для получения асимптотических выражений показателей надежности при высокой надежности всех элементов можно эффективно использовать теорему Реньи [5].

Для простоты рассмотрим дублированную систему. Пусть средняя наработка на отказ существенно больше среднего времени восстановления, т. е. $T\gg\tau$. В этом случае процесс функционирования элемента с восстановлением можно условно представить не в виде альтернирующего процесса восстановления, в котором чередуются периоды работоспособности и отказа, а просто в виде обычного процесса восстановления, в котором *относительно* кратковременные отказы представляются «точечными» событиями. Вероятность разрежения этого исходного потока событий равна вероятности того, что КПУ успешно переключит на место отказавшего основного элемента резервный, причем за время восстановления отказавшего элемента не произойдет отказа того, который в данное время *является основным.*

Рассмотрим дублированную систему с нагруженным резервом.

1. Первый тип КПУ (см. § 2.3), первая схема восстановления КПУ. Пусть для простоты $q_i=q$ для всех $i=1,2,\ldots$. Если вероятность несрабатывания КПУ мала, то в соответствии с теоремой

Реньи интенсивность потока отказов системы из-за невозможности по причине отказа КПУ переключиться на резервный элемент при отказе основного приближенно равна $\lambda^{(1)}=q/T$. Если время восстановления КПУ имеет распределение $G_{\text{КПУ}}(t)$, то в среднем после такого отказа система будет находиться в состоянии простоя в течение времени

$$\tau_{\Sigma}^{(1)}=\tau_{\text{КПУ}}=\int_0^{\infty} t\,dG_{\text{КПУ}}(t),$$

т. е. до восстановления КПУ.

Кроме того, дублированная система может отказать из-за того, что в момент отказа основного элемента резервный элемент, находившийся в нагруженном режиме, также находился в состоянии отказа и восстанавливался. Вероятность того, что отказ основного элемента приведет к отказу системы в целом,

$$k_{\text{р}}=\tau_{\text{р}}/(T_{\text{р}}+\tau_{\text{р}}),$$

где $\tau_{\text{р}}$ — время восстановления резервного элемента; $T_{\text{р}}$ — средняя наработка на отказ резервного элемента в нагруженном режиме.

Таким образом, интенсивность отказов дублированной системы по этой причине

$$\lambda^2=\lambda k_{\text{р}}.$$

Время простоя системы в этом случае зависит от нескольких факторов: от распределения времени восстановления отказавшего элемента, от числа ремонтных органов, дисциплины восстановления и других факторов, которые мы здесь рассматривать не будем. Однако предложенная методика легко позволит найти соответствующие характеристики в каждом конкретном случае.

Пусть восстановление неограниченное, т. е. при отказе двух элементов каждый из них восстанавливается независимо. В этом случае

$$\tau_{\Sigma}^{(2)}=\int_0^{\infty}\mathcal{P}\{\min(\eta,\tilde{\eta}_{\text{р}})\geqslant x\}\,dx, \tag{2.11}$$

где η — случайное время восстановления основного элемента; $\tilde{\eta}_{\text{р}}$ — остаточное случайное время восстановления резервного элемента.

Остаточное время $\tilde{\eta}_{\text{р}}$ имеет распределение

$$\gamma_{\text{р}}(t)=P\{\tilde{\eta}_{\text{р}}\leqslant t\}=1-\frac{1}{\tau_{\text{р}}}\int_t^{\infty}[1-G_{\text{р}}(x)]\,dx,$$

где $G_{\text{р}}(x)$ — распределение случайного времени восстановления $\eta_{\text{р}}$ резервного элемента; $\tau_{\text{р}}$ — среднее значение $\eta_{\text{р}}$.

Формулу (2.11) можно переписать в виде

$$\tau_{\Sigma}^{(2)}=\int_0^{\infty}[1-\gamma_{\text{р}}(t)]\,[1-G(t)]\,dt.$$

Из литературы известно (см., например, [5]), что для резервированной системы с восстановлением, состоящей из n независимых элементов, среднее время восстановления можно найти по формуле

$$\tau_{\Sigma}^{(n)}=\left(\sum_{1\leqslant i\leqslant n}\tau_i^{-1}\right)^{-1}.$$

Эта формула справедлива для случая, когда характеристики элементов зависят только от их собственного номера, но не зависят от той позиции, которую занимает в данный момент времени конкретный элемент.

В случае ограниченного восстановления дисциплина восстановления играет уже большую роль. В данном случае существенно, можно ли прерывать восстановление ремонтируемого резервного элемента, если известно, что восстановление основного элемента в среднем длится меньшее время (например, если во время отказа основного элемента возможна более достоверная и более точная аппаратная и программная диагностика места отказа).

При ограниченном восстановлении вычисление среднего времени простоя системы не вызывает трудности для каждого конкретного случая, поэтому мы на этих формулах останавливаться не будем.

Итак, суммарная интенсивность отказов системы равна сумме интенсивностей отказов по каждой из рассмотренных причин:

$$\lambda_{\Sigma}=\lambda^{(1)}+\lambda^{(2)},$$

а среднее время простоя системы определяется как средневзвешенное:

$$\tau_{\Sigma}=\frac{\tau_{\Sigma}^{(1)}\lambda^{(1)}+\tau_{\Sigma}^{(2)}\lambda^{(2)}}{\lambda^{(1)}+\lambda^{(2)}}.$$

2. Первый тип КПУ, вторая схема восстановления КПУ. Если КПУ проверяется с периодичностью θ, причем $\theta\ll T_{\text{КПУ}}$, то можно считать, что после возникновения отказа КПУ в течение в среднем времени $\theta/2$ КПУ находится в состоянии необнаруженного отказа, а затем еще в течение времени $\tau_{\text{КПУ}}$ восстанавливается. Если отказ основного элемента происходит в течение этого интервала времени, то это приводит к отказу системы.

Пропуская очевидные простые рассуждения, записываем

$$\lambda^{(1)}=\lambda q,$$

где

$$q=(0{,}5\theta+\tau_{\text{КПУ}})/(T_{\text{КПУ}}+0{,}5\theta+\tau_{\text{КПУ}}). \tag{2.12}$$

Однако естественно предположить, что, как только обнаруживается отказ КПУ из-за невозможности переключиться с отказавшего основного элемента на резервный, КПУ сразу же начинает восстанавливаться, поэтому

$$\tau_{\Sigma}^{(1)}=\tau_{\text{КПУ}}.$$

Выражения для $\lambda^{(2)}$ и $\tau_{\Sigma}^{(2)}$ остаются теми же, что и в первом случае.

3. Первый тип КПУ, третья схема восстановления КПУ. Этот вариант соответствует по существу непрерывному контролю КПУ. В этом случае в отличие от предыдущего нужно лишь в выражении (2.12) положить $\theta=0$.

4. Второй тип КПУ, первая схема восстановления КПУ. В этом случае благополучные переключения с отказавшего основного элемента на резервный осуществляются до тех пор, пока КПУ работоспособно. Однако после любого отказа КПУ первая же попытка переключения оказывается неудачной и начинается восстановление КПУ. В системах с таким режимом работы КПУ резервирование имеет смысл лишь тогда, когда $T_{КПУ} \gg T$. Будем условно считать, что после отказа КПУ основной элемент работает до своего отказа (а следовательно, и до отказа системы) в течение остаточного времени ζ. После этого наступает отказ системы на время восстановления КПУ. Таким образом, средняя наработка системы до отказа

$$T_{\Sigma}^{1} = T_{КПУ} + M\zeta. \qquad (2.13)$$

Если распределение случайной наработки основного элемента близко к экспоненциальному, то

$$M\zeta = T,$$

а если близко к вырожденному, то

$$M\zeta = 0{,}5T.$$

Конечно, нужно иметь в виду, что вырожденного распределения случайной наработки не бывает и возможен лишь случай, когда наработка имеет распределение с относительно малым коэффициентом вариации.

Понятно, что для данной схемы $\tau_{\Sigma}^{(1)} = \tau_{КПУ}$, а $\lambda^{(2)}$ и $\tau_{\Sigma}^{(2)}$ определяются, как и ранее.

5. Второй тип КПУ, вторая схема восстановления КПУ. В этом случае следует иметь в виду, что проверка работоспособности КПУ должна проводиться периодически через интервалы θ, которые существенно меньше средней наработки КПУ. Тогда, предположив, что проверки проводятся через неслучайное время, можно считать, что

$$\lambda^{(1)} = \lambda q,$$

где q определяется по формуле

$$q = (0{,}5\theta + \tau_{КПУ})/(T_{КПУ} + 0{,}5\theta + \tau_{КПУ}),$$

смысл которой очевиден.

Среднее время простоя системы

$$\tau_{\Sigma}^{(1)} = \tau_{КПУ},$$

поскольку естественно предположить, что при неудачной попытке переключения с отказавшего основного элемента на резервный восстановление КПУ осуществляется сразу же.

И в этом случае $\lambda^{(2)}$ и $\tau_{\Sigma}^{(2)}$ остаются прежними.

6. Второй тип КПУ, третья схема восстановления КПУ. Остаются в силе результаты предыдущего пункта, для которых следует лишь положить $\theta=0$.

2.9. О расчете надежности восстанавливаемых систем при произвольных распределениях

Предположение об экспоненциальности распределения наработки до отказа на практике выполняется далеко не для всех типов систем и элементов. Так, большинство механических блоков (принтеры, дисководы, пульты и т. п.) подвержено различным процессам износа и накопления повреждений, некоторые элементы (диски, магнитные носители других видов) характеризуются явно выраженным свойством старения во времени. Это приводит к необходимости исследования восстанавливаемых систем при очень общих предположениях.

С одной стороны, общие методы исследования таких систем весьма неконструктивны и громоздки. С другой, сами исходные статистические данные, используемые при расчетах надежности, весьма недостоверны. Иными словами, перед инженером возникает сложная задача: нужно проверить результаты расчетов, полученных на основании громоздких и порой трудно интерпретируемых математических моделей при наполнении их в определенном смысле сомнительными исходными данными.

Чаще всего возникающие при этом трудности преодолеваются использованием метода статистического моделирования, причем исходные данные берутся в «натуральном» виде, например в виде гистограмм, т. е. без учета каких-либо гипотез о виде распределений и без соответствующих аппроксимаций. Тем не менее ниже будут представлены некоторые простейшие методы расчета и приведены соображения по поводу развития тех или иных подходов к решению подобных задач.

Ненагруженный резерв. Этот случай является, пожалуй, единственным, когда удается провести расчет показателей надежности, а также получить простые асимптотические результаты. Используемый метод практически полностью совпадает с тем, который был рассмотрен при анализе дублированных систем для экспоненциального распределения наработки до отказа.

В данном случае цикл «работа—восстановление» состоит из двух отрезков $\xi+\eta'$, если на этом цикле отказа не произошло, и из двух отрезков $\xi+\eta''$, если на втором из них произошел отказ системы. Здесь, как и ранее, ξ — случайная наработка до отказа основного элемента; η' — случайное время восстановления при условии, что отказ системы не произошел за это время; η'' — случайное время работы системы до отказа всех резервных элементов при условии, что отказ происходит до окончания процесса восстановления. Каждый из этих отрезков имеет произвольное распределение. Дальнейший анализ даже в этом простом случае относительно несложен лишь тогда, когда общее время восстановления всей резервной группы будет существенно меньше в среднем наработки *основного элемента. При этом* если отказ системы не произошел, то можно считать, что к моменту предыдущего отказа основного элемента все резервные элементы, находящиеся в ненагруженном резерве, будут полностью восстановленными.

В этом случае можно, используя методы, изложенные в [9], записать преобразование Лапласа—Стилтьеса для распределения наработки системы до отказа, учитывая, что до этого события проходит случайное число циклов «работа—восстановление», причем распределение этого случайного числа является геометрическим:

$$\varphi(s) = \sum_{0 \leqslant k < \infty} qp^k [\varphi_0(s)\varphi'(s)]^k \varphi_0(s)\varphi''(s) = \frac{q\varphi_0(s)\varphi''(s)}{1 - p\varphi_0(s)\varphi'(s)},$$

здесь p — вероятность того, что на цикле «работа—восстановление» не произойдет отказа системы; $q=1-p$; $\varphi_0(s)$, $\varphi'(s)$ и $\varphi''(s)$ — преобразование Лапласа—Стилтьеса распределений случайных величин ξ, η' и η'' соответственно.

Обращение полученного преобразования (учитывая к тому же достаточно сложные выражения для преобразований Лапласа—Стилтьеса указанных распределений) представляет собой в общем случае специальную вычислительную задачу. Однако и здесь удается получить вполне инженерные формулы для высоконадежной резервируемой системы, когда $q \ll 1$, а время восстановления не имеет слишком большого коэффициента *вариации*. Если для безусловного времени восстановления η выполняется условие $M\eta \ll M\xi$, то, используя теорему Реньи, нетрудно для распределения наработки до отказа резервируемой системы записать выражение

$$P(t) \approx \exp(-q(M\xi)^{-1}t).$$

Нагруженный резерв. Более или менее строгое рассмотрение этого случая приводит к крайне громоздким и совершенно неконструктивным с вычислительной точки зрения результатам. В общих словах дело заключается в том, что математической моделью функционирования *такой системы является* суперпозиция нескольких процессов восстановления (в случае независимого восстановления каждого из элементов) либо еще более сложный по структуре случайный процесс. Такой процесс не имеет точек регенерации, т. е. его нельзя разделить на некоторые циклы, случайные процессы внутри которых являлись бы стохастически независимыми.

Однако могут быть высказаны полезные эвристические соображения относительно расчета показателей надежности для стационарного (установившегося) процесса функционирования. (Не будем останавливаться на детальном объяснении, когда процесс становится *стационарным.* Заметим лишь, что для большинства практических задач такие характеристики надежности можно вычислять по истечении 5—10 отказов элементов в системе, если коэффициент вариации распределения наработки до отказа не слишком близок к нулю.) При стационарном процессе функционирования и при независимом развитии процессов отказа и восстановления отдельных элементов системы можно считать, что к любому рассматриваемому моменту времени каждый из элементов характеризуется распределением остаточного времени жизни. Известно, что если случайный рекуррентный процесс образован случайными интервалами с распределением $F(x)$ со средним T, то стационарное остаточное время жизни имеет распределение

$$\widetilde{F}(x) = 1 - \frac{1}{T}\int_0^\infty [1 - F(x)]\,dx.$$

Далее, стационарный суммарный поток отказов элементов после «сильного перемешивания» с достаточным приближением можно считать близким к пуассоновскому, если число образующих его потоков достаточно велико. (На практике для прикидочных расчетов достаточно, чтобы число таких потоков было порядка 5—10.) В этом результирующем потоке отказы i-го элемента образуют подпоток с интенсивностью $\lambda_i = (T_i)^{-1}$. Отказ i-го элемента устраняется за случайное время с распределением $G_i(x)$. Отказ системы при отказе i-го элемента может наступить с вероятностью

$$q_i \simeq \int_0^\infty \prod_{k \neq i} q_k(t)\,dG_i(t) \simeq \tau_i \sum_{k \neq i}^{0} \lambda_k.$$

Итак, если выполняется условие $\max q_i \ll 1$, то можно записать достаточно простое приближенное выражение для вероятности безотказной работы рассматриваемой резервируемой системы $P(t) \approx \exp(-\sum_i \lambda_i q_i t)$ и для средней наработки

$$T \approx \left(\sum_i \lambda_i q_i\right)^{-1}.$$

Последовательная система. Возникающие в этом случае трудности расчета полностью совпадают с теми, которые встречались для системы с нагруженным резервом. Однако для высоконадежной системы здесь удается получить некоторые характеристики даже при исследовании нестационарного процесса функционирования. Для детального изучения этого случая мы отсылаем читателя к [7], где решение для независимого восстановления элементов дается через функции восстановления.

Для стационарного процесса функционирования последовательной системы, используя принципиально те же рассуждения, что и в предыдущем пункте, можно записать приближенное выражение для вероятности безотказной работы:

$$P(t) \approx \exp\left(-t\sum_k (T_k)^{-1}\right),$$

для средней наработки:

$$T=\left(\sum_k T_k^{-1}\right)^{-1}$$

и, наконец, для среднего времени простоя (восстановления) системы:

$$\tau=\sum_k\left(\int_0^\infty[1-G_k(x)]\,dx\right)T_k^{-1}\sum_i T_i^{-1},$$

что позволяет записать и достаточно простое выражение для коэффициента готовности:

$$K=\frac{1}{1+\tau T^{-1}}=\left[1+\sum_k T_k^{-1}\int_0^\infty[1-G_k(x)]\,dx\right]^{-1}.$$

При рассмотрении нестационарного режима можно предложить для малых значений t пренебрегать тем, что система является восстанавливаемой. Заметим, что если распределения таковы, что имеют производные не ниже k-го порядка при $t=0$ $(F_i^{(k)}=\alpha_i)$, то для расчета вероятности безотказной работы можно использовать распределение Гнеденко—Вейбулла

$$P(t)\approx\exp\left(-t^k\sum_i\alpha_i\right).$$

2.10. Информационные системы с ветвящейся структурой

Многие информационные системы имеют структуру специфического вида: информация от некоторого основного управляющего элемента (1-го уровня) поступает на несколько нижестоящих элементов (2-го уровня), каждый из которых передает, в свою очередь, информацию на элементы следующего уровня, и т. д. Это имеет место в системах, в которых осуществляется циркулярная передача информационных сигналов (или сигналов управления) либо сбор информации с нижестоящих элементов.

Похожую структуру могут иметь вычислительные системы, у которых основная мощная ЭВМ работает на систему ЭВМ меньшей мощности, каждая из которых, в свою очередь, снабжена разветвленной системой терминалов. Управление предполагается в соответствии со строгой иерархией подчиненности, причем исполнительный элемент системы считается функционирующим нормально, если он не только сам исправен, но и имеется связь его с основным (центральным) управляющим элементом (рис. 2.5).

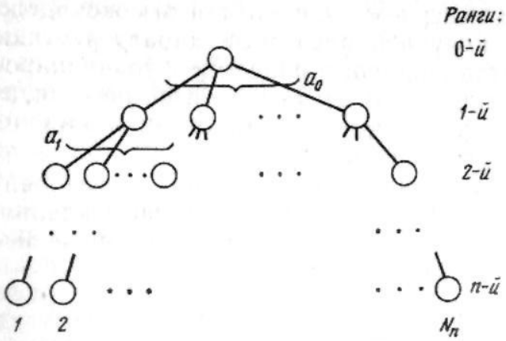

Рис. 2.5. Схема системы с симметричной ветвящейся структурой при прямом подчинении элементов

В соответствии со структурой симметричной ветвящейся системы элемент каждого i-го уровня (кроме последнего) имеет в своем непосредственном подчинении a_{i+1} элементов нижестоящего уровня. Будем считать, что полезный эффект создается только исполнительным элементом, что его нормальное функционирование обеспечивается всей цепочкой вышестоящих элементов, связывающих его с элементом 0-го ранга. В такой сложной системе вряд ли естественным будет деление всех состояний всего на два класса — работоспособности и отказа. Действительно, можно ввести критерий отказа: если число отказавших элементов n-го ранга больше некоторого ν, то система считается отказавшей. Однако резонно возникает вопрос: почему при ν—1 отказах элементов n-го ранга система была полностью работоспособна, а при ν отказах стала полностью неработоспособной? Естественнее считать, что эффективность функционирования постепенно ухудшается по мере роста числа отказавших исполнительных элементов. Пусть выходной эффект системы $E(z)$ зависит только от числа z нормально функционирующих выходных исполнительных элементов [10 [1]]:

$$E(z)=az+bz^2+cz^3+\dots.$$

Понятно, что в принципе полиномиальная форма позволяет достаточно хорошо аппроксимировать и пороговую функцию, задающую критерий отказа. Введем следующие определения и обозначения: число элементов, подчиненных одному элементу ближайшего i-го управляющего ранга, назовем его коэффициентом разветвления a_i, p_i — вероятность работоспособного состояния элемента i-го ранга, N_i — полное число элементов i-го ранга. Структуру будем полагать симметричной, т. е. все элементы одного и того же ранга идентичны по всем характеристикам. Предположим, что эффективность системы не зависит от взаимного расположения нормально функционирующих выходных элементов.

Эффективность функционирования системы будем оценивать математическим ожиданием ее выходного эффекта [10], т. е.

$$E=\sum_{0\leqslant z\leqslant N_n}P_n(z)\,E(z),$$

где $P_n(z)$ — вероятность того, что в системе работает z исполнительных элементов из общего числа N_n, или в другом виде

$$E=\mathrm{M}E(z)=a\mathrm{M}_1+b\mathrm{M}_2+c\mathrm{M}_3,$$

[1] См. также: **Ушаков И. А., Коненков Ю. К.** Оценка эффективности сложных ветвящихся систем с учетом надежности // Кибернетику — на службу коммунизму. — Т. 2 / Под ред. А. И. Берга, Н. Г. Бруевича, Б. В. Гнеденко. — М.: Энергия, 1964.

где M_k — начальный момент k-го порядка распределения случайной величины z.

Итак, для оценки эффективности рассматриваемой системы нужно знать начальные моменты распределения величины z. Для их определения запишем моментную производящую функцию для распределения числа нормально функционирующих элементов, подчиненных одному элементу $(n-1)$-го ранга. Для биномиального распределения эта производящая функция, как известно, имеет вид

$$\varphi(e^s) = (p_n e^s + q_n)^{a_n}.$$

Рассмотрим ситуацию, когда в системе нормально функционируют ν элементов $(n-1)$-го ранга. Тогда производящая функция числа нормально функционирующих исполнительных элементов равна произведению ν функций $\varphi(e^s)$. Вероятность такой ситуации определяется по формуле биномиального распределения и равна

$$C^{\nu}_{N_{n-1}} p^{\nu}_{N-1} q^{N_{n-1}-\nu}_{n-1}.$$

Моментная производящая функция для n-го ранга

$$\varphi_n(e^s) = \sum_{0\leqslant\nu\leqslant N_{n-1}} [\varphi(e^s)]^{\nu} C^{\nu}_{N_{n-1}} p^{\nu}_{n-1} q^{N_{n-1}-\nu}_{n-1} =$$

$$= (p_{n-1}\varphi(e^s) + q_{n-1})^{N_{n-1}} = \varphi_{n-1}(\varphi(e^s)),$$

где $\varphi(e^s)$ определено выше.

Таким образом, получена рекуррентная формула для моментов производящих функций, которая позволяет записать следующие рекуррентные выражения для двух первых начальных моментов:

$$M_n^1 = \left.\frac{\partial\varphi_n(s)}{\partial s}\right|_{s=0} = M^1_{n-1} a_n p_n,$$

$$M_n^2 = \left.\frac{\partial^2\varphi_n(s)}{\partial s^2}\right|_{s=0} = M^2_{n-1}(a_n p_n)^2 + M^1_{n-1} p_n q_n a_n,$$

или окончательно в замкнутой форме

$$M_n^1 = p_0 \prod_{i=1}^{n} a_i p_i,$$

$$M_n^2 = p_0 \prod_{1\leqslant i\leqslant n} a_i p_i \left(\prod_{1\leqslant i\leqslant n} a_i p_i + \sum_{1\leqslant i<n} q_i \prod_{i+1\leqslant k\leqslant n} a_k p_k \right).$$

Моменты более высоких порядков можно определить также без труда, однако они имеют достаточно громоздкие выражения и здесь не приводятся.

Для несимметричных ветвящихся систем удается построить доступную для исследования модель оценки эффективности функционирования лишь для случая, когда выходной эффект системы есть линейная функция числа работоспособных элементов.

Пусть ветвящаяся система имеет произвольную структуру, ее выходной эффект зависит, как и ранее, от нормального функционирования выходных элементов, причем элементы с разными номерами вносят различный вклад в общий выходной эффект системы. Обозначим через a_i долю выходного эффекта i-го выходного элемента системы, $i=1, ..., N_n$. По определению нормального функционирования выходного элемента требуется, чтобы он был работоспособен сам и при этом были бы исправны все элементы на предыдущих уровнях, которые осуществляют связь данного выходного элемента с основным управляющим элементом.

Выходные элементы взаимно зависимы, так как подчиняются одним и тем же управляющим элементам высшего уровня. В частности, все элементы системы могут одновременно прекратить нормальное функционирование, если откажет элемент верхнего уровня (0-го ранга).

Выходной эффект системы представляется суммой выходных эффектов отдельных исполнительных элементов системы. Следовательно, интересующее нас математическое ожидание выходного эффекта равно сумме математических ожиданий выходных эффектов отдельных элементов системы. Последнее утверждение следует из того, что математическое ожидание суммы случайных величин равняется сумме математических ожиданий даже в случае их зависимости.

Таким образом, если через p_i обозначить, как и ранее, вероятность исправного состояния i-го выходного элемента, а через p_{i1}, p_{i2} и т. д. вероятности работоспособного состояния i-го элемента 1-го ранга, 2-го ранга и т. д., которым подчинен данный элемент, то выходной эффект ветвящейся системы

$$E = p_0 \sum_{1\leqslant i\leqslant N_n} p_i a_i \prod_{1\leqslant j\leqslant n-1} p_{ij}.$$

Так же просто для линейной зависимости выходного эффекта от числа нормально функционирующих выходных элементов можно определить эффективность функционирования ветвящейся системы со сложным подчинением, т. е. когда возможны связи по

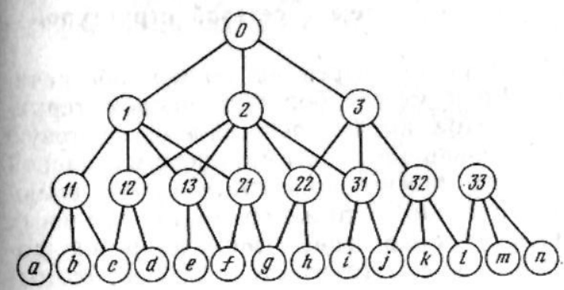

Рис. 2.6. Схема системы с ветвящейся структурой при сложном подчинении элементов

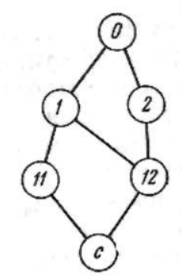

Рис. 2.7. Схема расчета вероятности нормального функционирования выходного элемента c

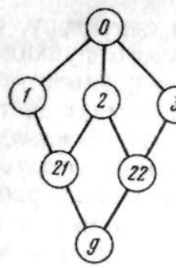

Рис. 2.8. Схема расчета надежности нормального функционирования выходного элемента g

управлению и через ранг и даже с несколькими управляющими элементами различных рангов. Для примера рассмотрим систему, представленную на рис. 2.6. Чтобы вычислить показатель эффективности функционирования такой системы, нужно для каждого выходного элемента составить отдельную схему расчета надежнсти, т. е. схему связи этого выходного элемента с элементом верхнего уровня.

Так, для выходного элемента c получаем схему расчета надежности, представленную на рис. 2.7. Формула расчета вероятности связности этого выходного элемента с элементом верхнего уровня имеет вид

$$r_c = p_c p_0 [p_1 p_2 (1 - q_{11} q_{12}) + q_1 p_2 p_{12} + p_1 q_2 (1 - q_{11} q_{12})] = p_c p_0 [p_1 (1 - q_{11} q_{12}) + q_1 p_2 p_{12}],$$

где $q = 1 - p$.

Приведем для примера еще схему расчета надежности для выходного элемента системы g (рис. 2.8) и запишем для него формулу для расчета вероятности связности его с элементом верхнего уровня:

$$r_g = p_0 p_g \{p_2 (1 - q_{21} q_{22}) + q_2 [1 - (1 - p_1 p_{21})(1 - p_3 p_{22})]\}$$

(при расчете фактически использован прием, основанный на разложении структуры относительно элемента 2).

Нахождение моментов распределения числа нормально функционирующих элементов более высокого порядка для ветвящихся систем со сложным подчинением существенно затрудняется даже для симметричных структур.

2.11. Показатели связности для систем с сетевой структурой

Большинство сетей связи и систем передачи данных, обеспечивающих информационный обмен между большим числом территориально разнесенных абонентов, имеют очень сложную сетевую структуру. Связь между отдельными пунктами информационной сети может осуществляться по многим возможным путям, включая транзит по целому ряду пунктов. В то же время для реальных сетей имеются и различные сложные ограничения, например допускается не более заданного число транзистов. Однако самым важным является то, что передача сообщения в сети занимает на определенное время те или иные канальные мощности, а также процессоры в узлах связи. В связи с этим более или менее адекватная модель сети связи — это сеть массового обслуживания. Более грубо сеть может быть описана при помощи потоковой модели. Чаще всего для получения различных показателей надежности и эффективности сетей связи используются статистические модели, реализуемые на современных быстродействующих ЭВМ.

На определенных этапах анализа функционирования сетей связи может оказаться полезным анализ связности любой выбранной пары абонентов или связности группы абонентов между собой или с группой других абонентов.

Ниже будет дан метод оценки связности двухполюсной сети при возможности отказов отдельных ребер. (Узлы предполагаются абсолютно надежными, хотя отсутствие этого предположения не приводит к принципиальным трудностям.)

Несколько идеализируя реальную информационную сеть, будем считать возможными любые транзитные пути передачи информации, не включающие циклов.

Строгий анализ информационной сети с произвольной структурой с целью определения вероятностных характеристик связности, по существу, возможен лишь методами, близкими к прямому перебору состояний. При этом каждое состояние анализируется в соответствии с выбранным критерием связности, что само по себе представляет достаточно сложную задачу. К тому же даже относительно простые реальные системы с числом элементов (каналов связи, узлов коммутации, абонентских пунктов) порядка 30—40 приводят к необходимости перебора миллионов состояний. Поэтому естественно, отказавшись от получения точных значений показателей, перейти к получению граничных (верхней и нижней) оценок вероятности связности.

Рассмотрим произвольную сложную сеть связи и оценим одну из важных характеристик — вероятность связи двух фиксированных абонентов. В качестве математической модели сети связи рассмотрим случайный граф: каналы связи — ребра, а узлы связи — вершины графа. Будем считать, что абоненты расположены непосредственно в узлах сети.

Введем в рассмотрение так называемую структурную функцию случайного двухполюсного графа (полюсами при данном анализе выбрана конкретная пара рассматриваемых абонентов):

$$\varphi(X) = \begin{cases} 1, & \text{если выбранные полюсы графа связаны друг с другом, т. е. если система работоспособна по критерию связности выбранной пары абонентов;} \\ 0 & \text{в противном случае,} \end{cases}$$

где $X = (x_1, \ldots, x_n)$ — вектор, компонентами которого являются индикаторы

$$x_i = \begin{cases} 1, & \text{если } i\text{-е ребро существует в графе;} \\ 0 & \text{в противном случае.} \end{cases}$$

(Для простоты будем считать, что все вершины в рассматриваемом случайном графе абсолютно надежны.)

Системы рассматриваемого типа называются системами с монотонной структурой, структурная функция которых обладает следующими естественными свойствами:

1) $\varphi(1, 1, ..., 1) = 1$, т. е. если все элементы системы работоспособны, то и вся система в целом работоспособна;

2) $\varphi(0, 0, ..., 0) = 0$, т. е. если все элементы системы отказали, то в состоянии отказа находится и вся система;

3) $\varphi(x_1, x_2, ..., x_n) \geqslant \varphi(x'_1, x'_2, ..., x'_n)$, если $x_i \geqslant x'_i$ для всех $i = 1, 2, ..., n$ (причем хотя бы для одного из индексов выполняется строгое неравенство), т. е. при отказе очередного элемента системы состояние системы может только ухудшиться (или, точнее, не улучшится);

4) замена любого элемента x_k на элемент x'_k не уменьшает вероятности связности (вероятности безотказной работы по критерию связи полюсов), если

$$p_k = \mathcal{P}\{x_k = 1\} = Mx_k \leqslant p'_k = \mathcal{P}\{x'_k = 1\} = Mx'_k.$$

Вернемся к двухполюсному графу, описывающему информационную сеть. Введем понятия простой цепи и простого разреза.

Простой цепью $\alpha_k(x)$ двухполюсного графа называется такое подмножество ребер этого графа, которое обеспечивает связность полюсов этого графа даже при отсутствии остальных ребер графа, но отказ любого из элементов из $d_k(X)$ при тех же условиях приводит уже к нарушению связности полюсов. Понятно, что $\alpha_k(X)$ есть последовательное соединение элементов со структурной функцией

$$\alpha_k(X) = x_1 \wedge x_2 \wedge ... \wedge x_{n_k} = \bigcap_{s \in a_k} x_s,$$

где n_k — число элементов в $\alpha_k(X)$; a_k — множество индексов элементов, входящих в $\alpha_k(X)$; $\wedge$ и $\bigcap$ — знаки логического умножения и произведения.

Понятно, что в общем случае в двухполюсном графе может быть несколько различных простых цепей.

Простым разрезом $\beta_j(X)$ двухполюсного графа называется такое подмножество ребер этого графа, отсутствие которых приводит к нарушению связности полюсов этого графа даже при наличии всех отдельных ребер графа, но введение в граф любого из элементов из $\beta_j(X)$ при тех же условиях приводит уже к наличию связности полюсов. Таким образом, $\beta_j(X)$ есть параллельное соединение элементов со структурной функцией

$$\beta_j(X) = x_1 \vee x_2 \vee ... \vee x_{m_j} = \bigcup_{s \in b_j} x_s,$$

где m_j — число элементов в $\beta_j(X)$; b_j — множество индексов элементов $\beta_j(X)$; $\vee$ и $\bigcup$ — знаки логического сложения и суммирования.

Структурная функция $\varphi(X)$ двухполюсного графа может быть выражена либо через структурные функции простых цепей

$$\varphi(X) = \bigcup_{1 \leqslant k \leqslant r} \alpha_k(X), \qquad (2.14)$$

либо через структурные функции простых разрезов

$$\varphi(X) = \bigcap_{1 \leqslant j \leqslant s} \beta_j(X), \qquad (2.15)$$

где r — полное число простых цепей; s — полное число простых разрезов сети по отношению к выбранной паре полюсов.

Оба эти выражения имеют простой содержательный смысл. Первое из них означает, что в двухполюсном графе для связи входа с выходом должна существовать хотя бы одна из простых цепей, а второе означает, что для связи входа с выходом в двухполюсном графе каждый из простых разрезов должен содержать хотя бы одно не отказавшее ребро (один работоспособный элемент), т. е. в графе не должно существовать ни одного разреза.

Иными словами, двухполюсный граф с любой структурой можно представить в виде параллельного соединения простых цепей или в виде последовательного соединения простых разрезов.

Для преобразования (2.14) и (2.15) к удобному виду используем теорему де Моргана

$$x \vee y = \overline{\bar{x} \wedge \bar{y}} \qquad (2.16)$$

или одну из ее эквивалентных форм, получающихся непосредственно из предыдущей при использовании свойства двойного отрицания для булевой переменной $(\bar{\bar{x}} = x)$:

$$\overline{\bar{x} \vee \bar{y}} = x \wedge y. \qquad (2.17)$$

Выражения (2.16), (2.17) можно легко обобщить на любое число переменных:

$$\bigcup_{1 \leqslant i \leqslant n} x_i = \overline{\bigcap_{1 \leqslant i \leqslant n} \bar{x}_i} \quad \text{и} \quad \bigcap_{1 \leqslant i \leqslant n} x_i = \overline{\bigcup_{1 \leqslant i \leqslant n} \bar{x}_i}.$$

Тогда для структурной функции разреза можно записать

$$\beta_j(X) = \bigcup_{s \in b_j} x_i = \overline{\bigcap_{s \in b_j} \bar{x}_s} = 1 - \bigcap_{s \in b_j} \bar{x}_s, \qquad (2.18)$$

а для параллельного соединения цепей

$$\bigcup_{1 \leqslant j \leqslant r} \alpha_k(X) = \overline{\bigcap_{1 \leqslant j \leqslant r} \overline{\alpha_k(X)}} = 1 - \bigcap_{1 \leqslant j \leqslant r} \overline{\alpha_k(X)}. \qquad (2.19)$$

Итак, с учетом (2.18) и (2.19) формулы (2.14) и (2.15) можно записать в виде

$$\varphi(X) = 1 - \bigcap_{1 \leqslant k \leqslant r} \Big(1 - \bigcap_{s \in a_k} x_s\Big),$$

$$\varphi(X) = \bigcap_{1 \leqslant j \leqslant s} \Big(1 - \bigcap_{s \in b_j} (1 - x_s)\Big).$$

Смысл этих структурных функций поясним на примере самой простой структуры, которая не может быть представлена в виде

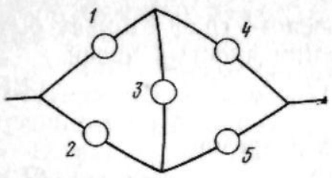

Рис. 2.9. Схема системы с мостиковой структурой

обычного параллельно-последовательного или последовательно-параллельного соединений, — на примере мостиковой схемы (рис. 2.9). Простыми цепями для этой структуры будут следующие подмножества элементов: (1, 4), (2, 5), (1, 3, 5), (2, 3, 4), а простыми разрезами — подмножества: (1, 2), (4, 5), (1, 3, 5), (4, 3, 2). Таким образом, структурную функцию можно записать в двух эквивалентных формах:

$$\varphi(X)=1-(1-x_1x_4)(1-x_2x_5)(1-x_1x_3x_5)(1-x_2x_3x_4),$$

$$\varphi(X)=(1-(1-x_1)(1-x_2))(1-(1-x_4)\times$$
$$\times(1-x_5))(1-(1-x_1)(1-x_3)).$$

Заметим, что единственную простую цепь имеет лишь последовательная система, и эта цепь состоит из всех n элементов системы.

Параллельная система имеет n простых цепей, каждая из которых состоит всего из одного элемента. В то же время параллельная система имеет единственный разрез, состоящий из всех элементов системы, а последовательная система имеет число разрезов, совпадающее с числом элементов, поскольку отказ каждого из них нарушает связь входа с выходом у этого простейшего двухполюсного графа.

Из перечисления простых цепей и простых разрезов для мостиковой схемы видно, что в различные цепи (аналогично и в разрезе) входят одни и те же элементы, т. е. эти цепи (разрезы) являются зависимыми. Иными словами, одноименные элементы в различных цепях (или разрезах) должны одновременно находиться либо в состоянии работоспособности, либо в состоянии отказа. Представление же в виде параллельного соединения простых цепей или в виде последовательного соединения простых разрезов из независимых элементов приводит к искажению истинного значения вероятности связности. Чтобы оценить, какое из представлений дает нижнюю, а какое верхнюю оценку, вспомним следующий известный из теории вероятностей факт: если события A и B положительно коррелированы, то

$$\mathcal{P}\{A\wedge B\}>\mathcal{P}\{A\}\mathcal{P}\{B\}.$$

При нескольких событиях имеет смысл говорить об их взаимной зависимости. Тогда для совокупности таких взаимно зависимых в совокупности событий можно записать неравенство вида

$$\mathcal{P}\{\bigcap_{1\leqslant i\leqslant n}A_i\}>\prod_{1\leqslant i\leqslant n}\mathcal{P}\{A_i\}.$$

Отсюда можно сразу написать для любого последовательного соединения простых разрезов неравенство

$$\mathcal{P}\{\bigcap_{1\leqslant j\leqslant r}\beta_j(X)=1\}>\prod_{1\leqslant j\leqslant r}\mathcal{P}\{\beta_j(X)=1\}$$

и для любого параллельного соединения простых цепей неравенство

$$\mathcal{P}\{\bigcup_{1\leqslant k\leqslant s}\alpha_k(X)=1\}=\mathcal{P}\{\overline{\bigcap_{1\leqslant k\leqslant s}\overline{\alpha_k(X)}}=1\}=$$

$$=1-\mathcal{P}\{\bigcap_{1\leqslant k\leqslant s}\overline{\alpha_k(X)}=1\}<1-\prod_{1\leqslant k\leqslant s}\mathcal{P}\{\overline{\alpha_k(X)}=1\}.$$

Отсюда следуют двусторонние оценки Эзари—Прошана[1] для вероятности связности двухполюсного графа [11, 12]:

$$\prod_{1\leqslant k\leqslant s}\left(1-\prod_{i\in b_k}q_i\right)\leqslant \mathrm{M}\varphi(X)=\mathcal{P}\{\varphi(X)=1\}\leqslant 1-\prod_{1\leqslant j\leqslant r}\left(1-\prod_{i\in a_j}p_i\right).$$

Для мостиковой схемы эти оценки имеют вид

$$(1-q_1q_4)(1-q_2q_5)(1-q_1q_3q_5)(1-q_2q_3q_4)\leqslant P\leqslant 1-(1-p_1p_2)\times$$
$$\times(1-p_4p_5)(1-p_1p_3p_5)(1-p_2p_3p_4).$$

Таким образом, представление произвольного двухполюсного графа в виде параллельного соединения простых цепей, составленных из независимых элементов, дает верхнюю оценку при вычислении вероятности связности, а представление в виде последовательного соединения простых разрезов, составленных также из независимых элементов, дает нижнюю оценку для той же вероятности.

Подобная двусторонняя оценка вероятности связности графа не является единственной. Другой способ построения таких оценок основывается на использовании только реберно-непересекающихся простых цепей и только реберно-непересекающихся простых разрезов (метод Литвака—Ушакова[2]). Понятно что для таких цепей и разрезов по определению выполняются условия

$$\beta_j(X)\cap\beta_l(X)=\varnothing,\ j\neq l;$$
$$\alpha_k(X)\cap d_s(X)=\varnothing,\ k\neq s.$$

Конструктивная процедура построения на основе исходного двухполюсного графа оценочного подграфа, состоящего из параллельно соединенных реберно-непересекающихся простых цепей, состоит в следующем. Выбираем произвольную простую цепь, например $\alpha_k(X)$. После исключения ребер этой цепи из исходного графа $\Gamma(X)$ получим подграф меньшей размерности $\Gamma_k(X)$. В полученном графе вновь выделяем простую цепь, которая по построению не будет иметь ни одного общего ребра с целью $\alpha_k(X)$. Подобная процедура продолжается до тех пор, пока получившийся на последнем шаге подграф не окажется несвязным либо пока не будут использованы при построении все ребра графа.

---

[1] Эти оценки называются так по имени авторов, получивших их впервые: **Эзари Дж., Прошан Ф.** Надежность связанных систем // Методы введения избыточности для вычислительных систем: Пер. с англ. / Под ред. В. С. Пугачева. — М.: Сов. радио, 1966.

[2] Эти оценки называются так по имени авторов, получивших их впервые в 1977 г. [13—15].

Понятно, что в общем случае возможно не единственное представление исходного графа через реберно-непересекающиеся простые цепи.

Конструктивная процедура построения оценочного графа, состоящего из последовательного соединения реберно-непересекающихся простых разрезов, заключается в следующем. Выбираем произвольный простой разрез. Все «левые» вершины, инцидентные данной совокупности ребер, стягиваем в одну вершину, а все «правые» вершины этой же совокупности ребер стягиваем в другую вершину. Это эквивалентно тому, что в исходный граф вводятся дополнительно «абсолютно надежные» ребра, если говорить об этой процедуре в терминах надежности. После этого образовывается новый граф, который состоит из трех частей: 1) «левого» фрагмента исходного графа, включающего в свой состав входную вершину двухполюсника; 2) последовательно включенного простого разреза, 3) «правого» фрагмента, включающего в свой состав выходную вершину.

Представление исходного графа в виде последовательного соединения реберно-непересекающихся простых разрезов также не единственно.

Примеры построения оценочных графов для мостиковой схемы с использованием реберно-непересекающихся цепей и реберно-непересекающихся разрезов приведены на рис. 2.10,*а*, *б*.

Из построения оценочного графа в виде последовательного соединения реберно-непересекающихся простых разрезов видно, что у любого оценочного графа вероятность связности выше, чем у исходного графа, поскольку при построении оценки в исходный граф вводятся условно «абсолютно надежные» ребра.

Аналогично можно показать, что оценочный граф в виде параллельного соединения реберно-непересекающихся простых цепей дает заниженную оценку для вероятности связности.

Если учесть неединственность представления исходного графа оценочными графами, то можно записать следующие итоговые оценки [16]:

$$\max_{1\leqslant r\leqslant N}\left\{1-\prod_{a_j\in\pi_r}\left(1-\prod_{i\in a_j}p_i\right)\right\}\leqslant M\varphi(X)\leqslant\min_{1\leqslant s\leqslant M}\prod_{b_j\in\sigma_s}\left[1-\prod_{i\in b_j}(1-p_i)\right],$$

где π_r — r-е подмножество реберно-непересекающихся простых цепей, образующих оценочный граф; σ_s — s-е подмножество реберно-непересекающихся простых разрезов, образующих оценочный граф; N — общее число оценочных графов, дающих нижнюю оценку вероятности связности исходного графа; M — общее число оценочных графов, дающих верхнюю оценку.

Для мостиковой схемы эти оценки примут вид

$$\max\{[1-(1-p_1p_2)(1-p_4p_5)],$$
$$p_1p_3p_5,\ p_2p_3p_4\}\leqslant P\leqslant\min\{(1-q_1q_4)(1-q_2q_5),$$
$$1-q_1q_3q_5,\ 1-q_2q_3q_4\}.$$

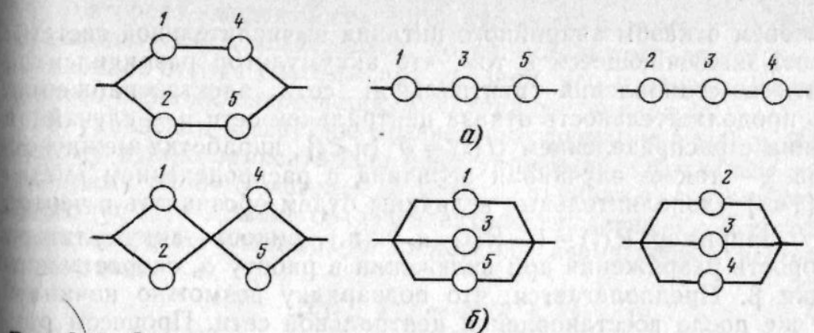

Рис. 2.10. Оценочные графы для мостиковой структуры, составленные из реберно-непересекающихся цепей и разрезов

Оценки Литвака—Ушакова являются в определенном смысле более конструктивными, чем оценки Эзари—Прошана, если говорить о чисто вычислительных аспектах, а также о точности оценок для сетей большой размерности. (Подробнее об этом см. в [16].)

2.12. Блокирование кратковременных отказов

В книге уже достаточно много внимания было уделено повышению надежности за счет введения структурного резервирования: резервных элементов, работающих в нагруженном режиме; запасных элементов, способствующих быстрому осуществлению восстановительных работ (что повышает показатели готовности); различного рода функционального резервирования (что повышает показатели безотказности) в виде формирования сетевых структур.

Однако этим не исчерпываются способы повышения надежности и эффективности функционирования систем. Существует еще один важный способ — использование имеющегося резерва времени при выполнении системой своих основных функций. Например, некоторые особо ответственные вычислительные системы, перерывы в электроснабжении которых чреваты очень серьезными последствиями, снабжаются автономными аккумуляторными источниками питания. Эти источники позволяют обеспечить нормальное функционирование вычислительной системы при отказах или авариях центральной сети электроснабжения в течение определенного (обычно сравнительно небольшого) периода времени. Однако за это время нормальное электропитание может быть восстановлено, и затем за определенное время аккумуляторный источник подзаряжается до номинального уровня. Если процесс подзарядки осуществляется со сравнительно небольшой скоростью, то с ощутимой вероятностью может возникнуть очередная аварийная ситуация, когда аккумулятор еще не подзарядился полностью.

Назовем отказом аварийного питания вычислительной системы событие, заключающееся в том, что аккумулятор разрядился до момента восстановления центральной сети электроснабжения. Пусть продолжительность отказа центральной сети η — случайная величина с распределением $G(t)=\mathcal{P}\{\eta\leqslant t\}$, наработка между отказами ξ — также случайная величина с распределением $R(t)=\mathcal{P}\{\xi\leqslant t\}$ (дополнительные величины будем обозначать с чертой сверху, например $\overline{R}(t)=1-R(t)$ и т. п.), емкость аккумулятора V, скорость разряжения при включении в работу α, скорость подзарядки β. Предполагается, что подзарядку возможно начинать сразу же после восстановления центральной сети. Процессы разряжения аккумулятора и его подзарядки для простоты будем считать линейными. (Экспоненциальный характер процесса незначительно усложнит методику.)

При этих условиях будем отыскивать вероятность безотказной работы вычислительной системы (при ее отказах только по причине исчерпания аккумуляторного питания).

Случайные процессы, аналогичные рассматриваемым, называются регенерирующими. Особенность этих процессов состоит в том, что для них характерны определенные состояния, после которых процесс развивается как бы заново. При анализе вероятности безотказной работы вычислительной системы с аккумулятором таким моментом регенерации удобно выбрать момент отказа центральной сети при условии, что к этому моменту аккумулятор был полностью подзаряжен.

Обозначим через θ случайную наработку до отказа системы, отсчитываемую именно от такого момента, когда емкость аккумулятора равна V и только что наступил отказ. Тогда случайную наработку системы до отказа можно записать как сумму двух случайных величин:

$$\theta^*=\zeta+\theta.$$

Таким образом, событие $\theta^*>t$ можно представить как реализацию одного из двух несовместных событий:

а) $\xi>t$,

б) $\xi=x$, где $x<t$, но при этом $\theta>t-x$.

Вероятность суммы несовместных событий равна сумме вероятностей этих событий. Это позволяет записать следующее интегральное уравнение, представляющее собой формулу полной вероятности для данного случая:

$$\overline{H}^*(t)=\overline{R}(t)+\int_0^t\overline{H}(t-x\mid V)\,dR(x), \tag{2.20}$$

где $R(t)=\mathcal{P}\{\xi\leqslant t\}$; $\overline{R}(t)=1-R(t)$; $H^*(t)=\mathcal{P}\{\theta^*\leqslant t\}$; $\overline{H}^*(t)=1-H^*(t)$; $H(t\mid V)=\mathcal{P}\{\theta\leqslant t\}$; $\overline{H}(t\mid V)=1-H(t\mid V)$.

Первое слагаемое в правой части уравнения (2.20) означает вероятность того, что за время t не произойдет отказа электросети, а второе — вероятность того, что если в момент $x\,(0\leqslant x\leqslant t)$ произойдет отказ электросети, то за оставшееся время $t-x$ не произойдет отказа системы при условии, что в начале рассматриваемого периода емкость аккумулятора составляла V.

В уравнении (2.20) две неизвестные функции $H^*(t)$ и $H(t\mid V)$, поэтому нужно записать еще одно уравнение. Обозначим случайное время работы системы до отказа, начиная с состояния аккумулятора V' в момент отказа электросети, через θ'. Для обозначения соответствующего распределения вероятностей используем запись вида

$$H(t\mid V')=\mathcal{P}\{\theta'<t\}.$$

Отличие $H(t\mid V')$ от $H(t\mid V)$ состоит только в отличии V' от $V: 0<V'<V$. Рис. 2.11 иллюстрирует процесс «расход — пополнение» емкости аккумулятора до отказа системы.

Запишем теперь уравнение для вероятности работы до отказа системы, начиная с момента начала отказа электросети, когда емкость аккумулятора равна V' (т. е. очередной отказ электросети наступил до момента подзарядки аккумулятора до исходного уровня V):

$$H(t\mid V')=\int_0^{\frac{V'}{\alpha}}\left[\int_0^{\frac{V-V'-\alpha x}{\beta}}H(t-x-y)\,\Big|\,(V'-\alpha x+\beta y)\,dR(y)+\right.$$
$$\left.+\overline{R}\left(\frac{V-V'-\alpha x}{\beta}\right)\overline{H^*}\left(t-x-\frac{V-V'-\alpha x}{\beta}\right)\right]dG(x) \tag{2.21}$$

при $\overline{H}(t\mid V')=1$ для $t\leqslant V'/\alpha$.

Поясним уравнение (2.21). Поскольку в момент отказа электросети емкость аккумулятора еще не успела восполниться до V, а достигла лишь $V'<V$, длительность очередного отказа не может превышать V'/α (смысл предела внешнего интеграла). Первое слагаемое в квадратных скобках означает, что после окончания отказа через время y возникает новый отказ, причем за это время не удалось восполнить емкость аккумулятора до уровня $V: V'-\alpha x+\beta y<V$. Второе слагаемое в квадратных скобках означает, что момент очередного отказа электросети происходит после момента полной подзарядки аккумулятора. Вероятность этого события есть $\overline{R}\left(\frac{V-V'-\alpha x}{\beta}\right)$. В этом случае с момента $x+$

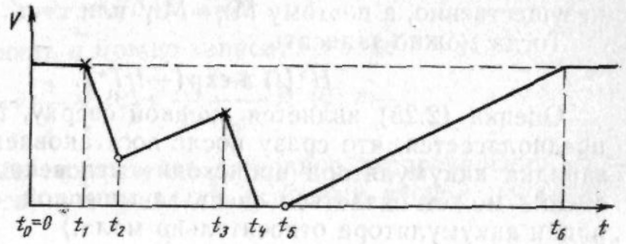

Рис. 2.11. График реализации расхода запаса

$+(V-V'-\alpha x)/\beta$, т. е. с момента полной подзарядки аккумулятора, вероятность безотказной работы системы будет равна $\overline{H}^*(t-x-(V-V'-\alpha x)/\beta)$.

Точное решение сформулированной задачи возможно только численными методами. Однако во многих практических задачах с достаточной для практики степенью точности могут быть использованы асимптотические методы. В данном случае такое приближенное решение возможно, если вероятность отказа системы при отказе электросети мала. Если исключить случаи паталогических распределений G и R (например, очень большие дисперсии), то достаточно потребовать выполнения условия $M\xi \gg V/\beta$.

Для приближенного определения вероятности безотказной работы системы воспользуемся известной в теории случайных процессов теоремой о разрежении случайных потоков (теоремой Реньи). Напомним, что в соответствии с этой теоремой любой рекуррентный случайный поток событий, образованный одинаково распределенными случайными величинами ξ, из которого исключаются события с вероятностью $1-q$, близкой к единице, с достаточной для практики точностью описывается пуассоновским потоком с параметром потока отказов $\lambda = q(M\xi)^{-1}$. На практике уже для $1-q \geqslant 0{,}9$ можно пользоваться пуассоновским приближением с достаточным основанием, так как неточность этой гипотезы с лихвой перекрывается неточностью исходных данных, а также различными идеализациями, используемыми при построении математической модели реального объекта. (Следует опять учесть, что дисперсия распределения случайной величины ξ не должна быть слишком большой.)

Заметим, что если ξ — экспоненциально распределенная случайная величина, то значение вероятности $1-q$ не играет роли: поток после разрежения остается пуассоновским.

В рассматриваемом случае

$$q = \overline{G}(V/\alpha).$$

Поэтому при $q \leqslant 0{,}1$ среднюю наработку до отказа системы можно оценить как

$$T^* \geqslant (T+\tau')/(\overline{G'}(V/\alpha)), \tag{2.22}$$

где T^*, T, τ' — средние значения величины θ^*, ξ, η' соответственно. Здесь η' обозначает случайную длительность отказа электросети при условии, что он не привел к отказу системы: $\eta' < V/\alpha$. Однако если $q \ll 1$, то можно считать, что усечение случайной величины несущественно, а поэтому $M\eta = M\eta'$ или $\tau = \tau'$.

Тогда можно записать

$$\overline{H}^*(t) \geqslant \exp(-t/T^*). \tag{2.23}$$

Оценка (2.23) является оценкой сверху, так как здесь неявно предполагается, что сразу после восстановления электросети подзарядка аккумулятора происходит мгновенно. (Заметим, что эта оценка может оказаться очень завышенной, если скорость подзарядки аккумулятора относительно мала.)

Оценку снизу получим, если сделаем следующее предположение: к отказу систему приводит и случай, когда с момента начала подзарядки аккумулятора происходит очередной отказ электросети через время, меньшее чем $\varkappa = V/\beta$.

Это означает, что при предыдущем отказе предполагается практически полная разрядка аккумулятора, а новый отказ электросети, возникший во время его подзарядки, заведомо окажется критическим для системы. Кроме того, конечно, отказ системы возникает и в том случае, если полная разрядка аккумулятора происходит сразу. Ясно, что такое огрубление приведет к завышению вероятности отказа системы.

Таким образом, отказ вычислительной системы при очередном отказе электросети может возникнуть с вероятностью, не большей

$$\overline{q} = \overline{G}\left(\frac{V}{\alpha}\right) + G\left(\frac{V}{\alpha}\right) R\left(\frac{V}{\beta}\right) = 1 - G\left(\frac{V}{\alpha}\right)\overline{R}\left(\frac{V}{\beta}\right).$$

Тогда выражение (2.22) примет вид

$$T^* \leqslant \frac{T+\tau}{1 - G\left(\frac{V}{\alpha}\right)\overline{R}\left(\frac{V}{\beta}\right)}.$$

(Заметим, что на практике дополнительно выполняется еще и условие $T \gg \tau$, т. е. величиной τ в числителе можно пренебречь.)

Оценку снизу можно улучшить следующим образом. (Заметим, что для практики оценка снизу важнее, так как дает гарантированную оценку вероятности.) Условие появления отказа электросети за время $\varkappa = V/\beta$ является слишком жестким в качестве критерия отказа, так как в среднем подзарядка аккумулятора осуществляется за меньшее (на практике даже за существенно меньшее) время. Действительно, вероятность появления очередного отказа электросети в течение подзарядки аккумулятора после предыдущего отказа определяется как

$$\tilde{q} = \int_0^{V/\alpha} R\left(\frac{\eta\alpha}{\beta}\right) dG(\eta).$$

Понятно, что $\tilde{q} < \overline{R}(V/\beta)$.

Если функция распределения наработки до отказа электросети разложима в ряд Тейлора в точке $t=0$, т. е.

$$R(t) = \frac{\alpha}{\beta} R'(0)\, t + \frac{\alpha^2}{2\beta^2} R''(0)\, t^2 + \ldots,$$

то искомую вероятность $\tilde{q}$ можно записать в виде

$$\tilde{q} = \frac{\alpha}{\beta} R'(0)\, \tau' + \frac{\alpha^2}{2\beta^2} R''(0)\, m_2,$$

где τ' — условное математическое ожидание распределения $G(x)$ при $\eta \leqslant V/\alpha$; m_2 — соответствующий условный второй начальный момент этого же распределения. Понятно, что и τ' и m_2 меньше, чем соответствующие величины, вычисленные без учета усечения распределения $G(x)$. В связи с этим

$$\tilde{q} < \frac{\alpha}{\beta} R'(0)\tau + \frac{\alpha^2}{2\beta^2} R''(0) m_2.$$

Если $R(t)$ — экспоненциальное распределение с параметром λ, то

$$\tilde{q} \leqslant \frac{\alpha}{\beta}\alpha\tau + 0,5\left(\frac{\alpha\lambda\tau}{\beta}\right)^2 + \ldots$$

Следовательно, можно записать уточненную верхнюю оценку

$$T^* \leqslant \frac{T+\tau}{1 - G\left(\frac{V}{\alpha}\right)\tilde{q}}.$$

Получение улучшенных верхней и нижней оценок $\overline{q}$ и $\underline{q}$ можно продолжить, если начать учитывать влияние более длинных цепочек «цепляющихся» друг за друга отказов. Заметим, что сама эта процедура может рассматриваться как вычислительная рекуррентная процедура решения системы уравнений (2.20), (2.21) для оговоренных выше соотношений параметров T, τ, α, β и V.

2.13. Решение вычислительной задачи с учетом простоев ЭВМ (без обесценивания)

Большие вычислительные задачи обычно формируются таким образом, чтобы отказ (или сбой) ЭВМ не приводил к необходимости полного повторного решения задачи. Для этого задача разделяется на отдельные фрагменты, после каждого из которых записываются промежуточные результаты решения. Это позволяет повторно решать лишь отдельные этапы задачи, на которых произошло нарушение вычислительного процесса. Таким образом, после каждого сбоя, если он обнаруживается мгновенно, приходится повторять решение лишь части программы от начала вычислительного фрагмента. Если сбой произошел через время ξ после начала счета последнего фрагмента, то это случайное время можно считать своеобразным простоем системы, так как это время, потерянное для вычислительного процесса. Если рассматриваются не сбои, а отказы, то кроме времени ξ в длительность простоя ЭВМ следует отнести еще и время восстановления. В этом случае можно считать, что задача длительности t_0 будет решена в течение времени θ, если суммарная наработка $\xi = \Sigma\xi_i$ за время θ будет не меньше чем t_0.

Впервые, видимо, данная задача была сформулирована в [9].

Рассмотрим альтернирующий процесс восстановления, т. е. случайный процесс, представляющий собой чередующуюся последовательность интервалов работоспособности и простоя. При этом предполагается, что все интервалы работоспособности одинаково

распределены с функцией распределения $F(t)$, а все интервалы простоя также одинаково распределены с функцией распределения $G(t)$, причем все рассматриваемые случайные величины взаимно независимы. Обозначим предварительно $G^{*k}(t)$ и $F^{*k}(t)$ k-кратные свертки соответствующих распределений.

Пусть наработка t_0 достигается именно на $(k+1)$-м периоде работы. Это означает, что

$$\sum_{1 \leqslant i \leqslant k} = \xi_i < t_0 \leqslant \sum_{1 \leqslant i \leqslant k+1} \xi_i.$$

При этом на интервале θ должно возникнуть ровно k отказов, которые в сумме продолжались в течение времени $\theta - t_0$. На интервале $\theta - t_0$ возникнет k или более отказов с вероятностью

$$P_k = G^{*k}(\theta - t_0).$$

Если записать аналогичное выражение для P_{k+1}, то можно заметить, что ровно k отказов происходит с вероятностью

$$p_k = P_{k+1} - P_k.$$

Таким образом, на отказе θ должно произойти ровно k отказов с суммарным простоем $\theta - t_0$, при этом k интервалов полезной наработки еще не достигли в сумме величины t_0 и отрезок ξ_{k+1} обязательно накрост точку θ, а суммарная наработка будет не менее t_0. В результате, поскольку k пробегает значения 0, 1, 2, ..., можно записать окончательную формулу для вероятности $R(\theta, t_0)$ того, что суммарная полезная наработка за время θ будет не меньше t_0 — достигнет величины s. Понятно, что

$$\mathcal{P}\{s_t < x\} = \mathcal{P}\{T_x > t\}.$$

Заметим, что если за время t суммарная наработка достигла величины x, то это означает, что суммарный простой за это же время достиг величины $t - x$. Понятно, что момент достижения заданной наработки приходится обязательно на интервал работоспособности. Это означает и то, что на рассматриваемом интервале времени число отказов будет равно числу восстановлений. Если обозначить через $\nu_1(x)$ число отказов, а через $\nu_2(x)$ число восстановлений за время x, то

$$\nu_1(x) = \nu_2(T_x - x). \tag{2.24}$$

Поскольку $T_x > t$,

$$\nu_1(x) \geqslant \nu_2(t - x).$$

Обозначим искомую вероятность достижения суммарной наработки не менее t_0 за время θ через $\Phi(\theta, t_0)$. Эту вероятность можно записать в виде

$$\Phi(\theta, t_0) = \sum_{0 \leqslant k < \infty} [G_k(\theta - t_0) - G_{k+1}(\theta - t_0)] F_k(t_0).$$

Для экспоненциальных распределений F и G с параметрами λ и μ соответственно

$$\Phi(\theta, t_0) = \sum_{0 \leqslant n < \infty} \frac{[M(\theta - t_0)]^n}{n!} e^{-\mu(\theta - t_0)} \sum_{n < k < \infty} \frac{(\lambda t_0)^k}{k!} e^{-\lambda t_0}.$$

При больших отношениях θ/t_0, как показано в [9], величина s_θ распределена асимптотически нормально со средним

$$\bar{s}_\theta = \frac{T}{T+\tau}\,\theta = K\theta,$$

где k — коэффициент готовности, и дисперсией

$$D = \frac{T^2\tau^2(\varkappa_1 + \varkappa_2)}{(T+\tau)^3}\,\theta = K^2(1-K)\,\tau\theta\,(\varkappa_1 + \varkappa_2),$$

где T — средняя наработка на отказ; τ — среднее время простоя; $\varkappa_1 = \sigma_T^2/T^2$, $\varkappa_2 = \sigma_\tau^2/\tau_2$ — коэффициенты вариации; σ_T^2 и σ_τ^2 — дисперсии соответствующих распределений.

В заключение заметим, что для экспоненциального распределения $\varkappa = 1$, и это приводит к незначительному упрощению выражения для D:

$$D = 2K^2(1-K)\,\tau\theta.$$

2.14. Решение вычислительной задачи с учетом сбоев ЭВМ (при полном обесценивании работы)

Можно рассмотреть следующую идеализированную схему вычислительного процесса: если в процессе решения некоторой задачи в вычислительной системе возникает сбой, то он сразу же обнаруживается и решение задачи возобновляется сначала. Возникает вопрос: с какой вероятностью может быть выполнена задача длительности t_0, если в вычислительной системе возникает поток сбоев, а полное время, отведенное на решение задачи равно θ, причем $\theta > t_0$.

Обозначим вероятность того, что задача длительности t_0 будет выполнена за время t, через $\Phi(t)$. Пусть сбои формируют рекуррентный поток, образованный одинаково распределенными независимыми случайными интервалами времени с распределением $H(t)$. В этом случае для любого t можно записать

$$\Phi(t) = [1 - H(t_0)]\,\delta_{t_0}(t) + \int_0^{t_0} \Phi(t-x)\,dH(x), \qquad (2.25)$$

где

$$\delta_{t_0}(t) = \begin{cases} 1, & \text{если} \quad t \leqslant t_0, \\ 0, & \text{если} \quad t > t_0 \end{cases}$$

— индикаторная функция.

Формула (2.25) имеет простой вероятностный смысл: для того, чтобы задача длительности t_0 была решена, необходимо, чтобы произошло одно из двух событий:

1) ни одного сбоя не будет на интервале $[0, t_0]$;

2) если первый сбой произойдет в момент $x < t_0$, то требуется, чтобы на оставшемся интервале $[x, t]$ задача была бы успешно решена (при этом предполагается, естественно, что если оставшееся время $t - x$ меньше требуемого t_0, то задача принципиально не может быть решена формально за счет индикаторной функции $\delta_{t_0}(t)$ соответствующая вероятность будет обращаться в нуль).

Уравнение (2.25) можно решить стандартным способом, используя, например, какую-нибудь итерационную вычислительную процедуру. Представляет интерес рассмотрение двух предельных случаев: 1) запас времени относительно невелик, например меньше удвоенного времени выполнения задачи; 2) запас времени очень велик, но велика и вероятность появления сбоя за время выполнения задачи $H(t_0)$.

В первом случае приближенное решение можно получить на основании следующих рассуждений. Задача может быть выполнена при отсутствии сбоев на начальном участке $[0, t_0]$. Если сбой (хотя бы один) все же возник, то задачу можно решить, если сбои будут на начальном интервале $[0, \theta - t_0]$. По формуле полной вероятности запишем

$$\Phi(\theta) \approx \overline{H}(t_0) + [1 - \overline{H}(\theta - t_0)]\,\overline{H}(t_0) = \overline{H}(t_0)\,[1 + H(\theta - t_0)].$$

Во втором случае можно воспользоваться результатом, известным из теории суммирования случайного числа случайных величин. Допустим, что до первого периода успешного выполнения задачи наступает достаточно большое число сбоев, прерывающих решение задачи и вынуждающих начинать вычислительный процесс заново. Вероятность того, что на очередном цикле задача не будет решена, равна $H(t_0)$.

При этом распределение случайного времени до прерывания отличается от распределения случайного времени между сбоями, так как первое всегда меньше t_0. Условное распределение интересующей нас величины

$$H^*(x) = \begin{cases} H(x)/H(t_0) & \text{при} \quad x < t_0, \\ 1 & \text{при} \quad x \geqslant t_0. \end{cases}$$

Отсюда математическое ожидание наработки до очередного сбоя вычислительного процесса при условии, что задачу не удается решить

$$T^* = \frac{1}{H(t_0)} \int_0^{t_0} x\,dH(x).$$

Таким образом, до наступления события, состоящего в решении вычислительной задачи длительности t_0, происходит случайное число прерываний, причем оно имеет геометрическое распределение.

Согласно тождеству Вальда

$$T = M \sum_{1 \leqslant i \leqslant \nu} \xi_i = M\nu M\xi,$$

где ν — случайное число слагаемых (число прерываний до успешного решения); ξ — случайное слагаемое (длительность работы до прерывания).

Если $M\nu$ велико, например порядка 10 и более, то случайная величина

$$\xi = \sum_{1 \leq i \leq \nu} \xi_i$$

имеет приближенно экспоненциальное распределение. (Это приближение следует из теоремы Реньи о разрежении рекуррентного потока.)

Далее

$$P\{\nu = k\} = [H(t_0)]^k \overline{H}(t_0).$$

Известно, что для геометрически распределенной величины ν

$$M\nu = [\overline{H}(t_0)]^{-1}.$$

Если сбои образуют пуассоновский поток с параметром λ, то

$$M\nu = e^{\lambda t_0},$$

$$T^* = (1 - e^{-\lambda t_0})^{-1} \int_0^{t_0} \lambda x e^{-\lambda x} dx = \lambda^{-1} (1 - e^{-\lambda t_0})^{-1} [1 - e^{-\lambda t_0} (1 + \lambda t_0)],$$

$$T = \lambda^{-1} (1 - e^{-\lambda t_0})^{-1} [1 - e^{-\lambda t_0} (1 + \lambda t_0)] e^{\lambda t_0}.$$

Таким образом, чтобы задача длительностью t_0 была бы решена в течение времени θ, нужно, чтобы случайная величина ξ была меньше $\theta - t_0$, поскольку последний интервал ξ_i должен быть больше t_0. Вероятность этого события

$$\Phi(\theta) \approx 1 - e^{-(\theta - t_0)/T}.$$

Более подробное изложение вопросов точного расчета показателей надежности систем с временны́м резервированием можно найти в [5][1].

2.15. Стареющий элемент

Иногда специалисту по надежности, которому предстоит рассчитать вероятностные характеристики проектируемой вычислительной системы, бывает известна лишь ограниченная статистика о надежности используемых элементов. (Напомним, что под элементом в широком смысле слова понимается любая часть системы, включая и любые подсистемы.) Малая статистика, как правило, имеющаяся на практике, позволяет говорить более или менее достоверно лишь об оценках средней наработки до отказа и в лучшем случае об оценках для дисперсии распределения наработки элемента. Если мы знаем лишь среднюю наработку, то, как известно, невозможно получить никакой нетривиальной информации о самом виде распределения наработки до отказа.

[1] См. также: **Черкесов Г. Н.** Надежность технических систем с временной избыточностью / Под ред. А. М. Половко. — М.: Сов. радио, 1974.

Принятие гипотезы об экспоненциальности распределения наработки до отказа не представляется неоспоримым.

Если известна дополнительно еще дисперсия распределения, то для произвольного распределения можно воспользоваться известной оценкой Чебышева. Однако эта оценка, являясь универсальной, имеет весьма ограниченное применение на практике, так как в общем случае определяет границы слишком грубо. Естественно ожидать сужения этих границ для оценки неизвестной функции распределения, если о ней известна какая-либо дополнительная информация.

Оказывается, что существенное улучшение оценок удается получить, если случайная величина ξ является «стареющей». Старение является достаточно понятным и естественным свойством. Оно означает, что вероятностные характеристики надежности некоторого изделия ухудшаются с течением времени. Например, старение некоторого изделия может означать возрастание с течением времени интенсивности отказов либо убывание среднего значения остаточного времени жизни в зависимости от длительности уже «прожитой жизни». Существуют и другие формальные определения старения, которые читатель может найти в [12].

В вычислительной технике наиболее явно свойствами старения характеризуются магнитные носители (ленты, диски), а также точная механика (лентопротяжки, дисководы). Отказы принтеров также обусловлены износом: правильно собранные и имеющие механические детали из хороших материалов, как правило, не отказывают на начальном периоде эксплуатации. К таким отказам можно отнести и отказы вентиляционной техники, холодильных установок.

Качественная на первый взгляд информация о старении приводит к возможности формулирования вполне строгих математических условий. Так, для «стареющих» случайных величин могут быть определены [17] соответствующие ограничения на все моменты распределения:

$$a_{n+1} a_{n-1} \leqslant a_n^2,$$

где $a_k = M_k / k!$; M_k—k-й начальный момент.

В частности, для первых двух известных моментов — математического ожидания и дисперсии — можно сформулировать условие и в иной форме: коэффициент вариации «стареющей» случайной величины не превышает единицы.

Зачастую суждение о том, является ли данное распределение стареющим или нет, можно сделать и без каких-либо конкретных статистических данных — просто на основании априорных качественных сведений и достаточно общих соображений. Например, если отказ возникает в результате износа или из-за накопления дефектов до некоторого критического уровня, то предположение о старении является вполне естественным: чем больше проходит времени с момента начала работы, тем с большей вероятностью возникновение очередного дефекта может привести к превышению этого критического уровня.

Вообще говоря, различные типы старения, о которых говорилось ранее, отличаются по своим свойствам. Поэтому для определенности будем говорить лишь о старении, характеризуемом возрастанием интенсивности отказов.

Напомним, что интенсивность отказов есть функция, определяемая как условная плотность распределения вида

$$\lambda(t) = \frac{f(t)}{P(t)} = \frac{-d}{dt} \ln P(t),$$

т. е. это условная плотность распределения момента отказа в момент t при условии, что до этого момента отказ не произошел. Это позволяет записать вероятность безотказной работы в виде

$$P(t) = \exp\left(-\int_0^t \lambda(x)\,dx\right),$$

или, если обозначить $\Lambda(t) = \int_0^t \lambda(x)\,dx$, то

$$P(t) = \exp(-\Lambda(t)).$$

Функция интенсивности $\lambda(t)$ рассматриваемого типа «стареющих» распределений является возрастающей (точнее, неубывающей). Тогда понятно, что функция $\Lambda(t)$ является возрастающей и выпуклой вниз функцией. Функция может иметь два типа поведения:

1) возрастая, стремится к некоторой линейно возрастающей функции $\lambda^* t + c$ (рис. 2.12, кривая *2*) это соответствует стремлению функции $\lambda(t)$ при возрастании к некоторой константе λ^* (рис. 2.13, кривая *2*);

2) возрастает так, что для всех t будет возрастать производная $\Lambda'(t)$ (рис. 2.12, кривая *1*); это соответствует неограниченному возрастанию самой функции $\lambda(t)$ (рис. 2.13, кривая *1*).

Видно, что только кривая 2 $\Lambda(t)$ не пересекает $\Lambda(t) = \lambda^* t$, т. е. это возможно только при ограниченном росте функции $\lambda(t)$. Во всех остальных случаях неограниченного возрастания $\lambda(t)$ лю-

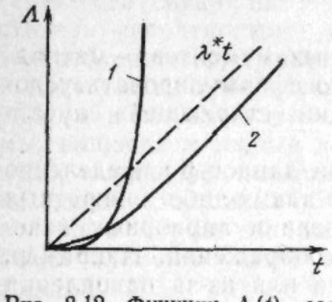

Рис. 2.12. Функции $\Lambda(t)$, соответствующие функциям $\lambda(t)$, представленным на рис. 2.13

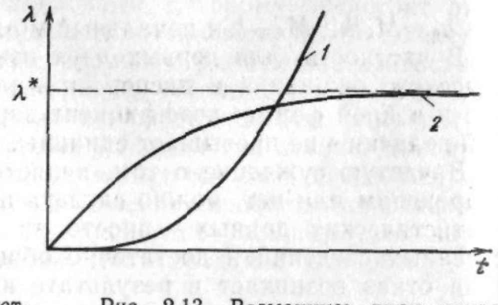

Рис. 2.13 Возможные типы непрерывных монотонных возрастающих функций интенсивности $\lambda(t)$

бая линейная функция $\lambda^* t$ пересекается функцией $\Lambda(t)$, соответствующей стареющему распределению. Эти предварительные результаты позволяют получить очень простые и изящные оценки [11, 12] неизвестной функции $P(t)$ для «стареющего» распределения, когда известно лишь математическое ожидание T соответствующей случайной величины (рис. 2.14)

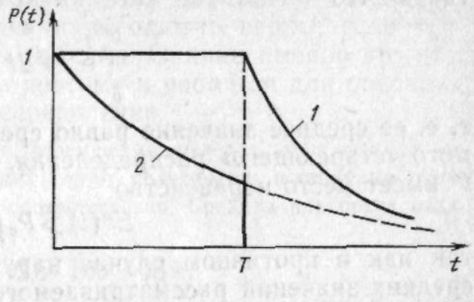

Рис. 2.14. Верхняя (*1*) и нижняя (*2*) границы для функции распределения стареющего элемента

$$a(t) \leqslant P(t) \leqslant A(t),$$

где $a(t)$ и $A(t)$ соответственно нижняя и верхняя граничные оценки, определяемые как

$$a(t) = \begin{cases} e^{-t/T} & \text{при } t < T, \quad (2.26) \\ 0 & \text{при } t \geqslant T, \quad (2.27) \end{cases}$$

$$A(t) = \begin{cases} 1 & \text{при } t \leqslant T, \quad (2.28) \\ e^{-\omega_t t} & \text{при } t > T; \quad (2.29) \end{cases}$$

ω_t, в свою очередь, определяется из условия

$$\int_0^t e^{-\omega_t t}\,dt = T$$

или, если его записать после интегрирования,

$$1 - \omega_t T = e^{-\omega_t t}.$$

Оценки (2.27) и (2.28) тривиальны и поэтому не требуют объяснения: они следуют из рис. 2.12, 2.13.

Оценку (2.26) можно пояснить следующим образом. Одним из предельных стареющих распределений является вырожденное распределение $D(t)$, т. е. распределение неслучайной постоянной величины:

$$D(t) = \begin{cases} 1 & \text{при } t \leqslant T, \\ 0 & \text{при } t > T. \end{cases}$$

Понятно, что до момента $t = T$ функция $D(t)$ будет проходить выше экспоненты $\exp(-t/T)$, что и соответствует нижней границе (2.26).

Оценку (2.29) можно истолковать следующим образом. Рассмотрим произвольный момент времени t^* и построим усеченную справа в этой точке экспоненту

$$E^*(t) = \begin{cases} e^{-\omega_t t} & \text{при } t \leqslant t^*, \\ 0 & \text{при } t > t^*, \end{cases}$$

такую, что

$$\int_0^{t^*} e^{-\omega_t t} dt = T,$$

т. е. ее среднее значение равно среднему значению рассматриваемого «стареющего» распределения. При таком построении в точке t^* имеет место неравенство

$$E^*(t) > P_c(t^*), \qquad (2.30)$$

так как в противном случае нарушалось бы условие равенства средних значений рассматриваемого «стареющего» распределения и усеченного экспоненциального распределения. Неравенство (2.30) и приводит к оценке (2.29).

Таким образом, на ранней стадии проектирования, когда мы не располагаем никакими достоверными данными о функции распределения наработки до отказа, кроме среднего значения и общих соображений о наличии феномена старения можно все же получить достаточно хорошие двусторонние оценки значений вероятности безотказной работы для заданных интервалов времени.

Заметим, что если дополнительно известна дисперсия неизвестного стареющего распределения, то могут быть получены гораздо более точные двусторонние оценки. (Дисперсия хорошо оценивается практически по той же исходной информации, что и среднее значение, если только последнее не вычисляется на основании известной суммарной наработки и общего числа отказов.) Построение этих оценок требует решения достаточно сложных систем трансцендентных уравнений. Для практического использования имеются соответствующие границы, протабулированные в широких пределах [11] и с необходимым для инженерных нужд шагом значений аргументов (в данном случае время, измеряемое в относительных единицах, и коэффициент вариации). Здесь же приведем одну очень простую и конструктивную приближенную оценку отклонения вероятности безотказной работы для «стареющего» распределения от вероятности для экспоненциального распределения

$$|P(t) - e^{-t/T}| \leqslant 1 - \sigma/T,$$

где σ — среднее квадратическое отклонение. Эта оценка хороша для коэффициентов вариации, близких к единице. Понятно, что для малых коэффициентов вариации она травиальна. Доказательство этой оценки весьма сложно и здесь не приводится.

У многих практиков имеется представление, что существуют только стареющие изделия. В основу такого представления кладутся примерно такие рассуждения: всякое изделие имеет свою вполне определенную, хотя и не известную, наработку от отказа, поэтому если какое-то время мы уже пронаблюдали за ним и оно не отказало, то до отказа времени осталось меньше, чем было до начала наблюдений. Такие рассуждения связаны с определенным непониманием сути вероятностной природы случайных явлений, в том числе случайных величин типа наработки до отказа.

Приведенные выше рассуждения абсолютно верны, если точно известно, каков срок службы изделия. Однако именно это нам и не известно априори и именно поэтому и вводятся для описания случайных величин функции распределения.

Приведем простой пример. Пусть перемешаны изделия двух равных по объему партий. Изделия каждой из них имеют практически постоянные наработки до отказа, скажем 10 и 100 ч соответственно. Средняя наработка изделия смешанной общей группы изделий

$$T = 0{,}5 \cdot 10 + 0{,}5 \cdot 100 = 55 \text{ ч}.$$

Допустим, что наугад выбранное изделие проработало в течение 20 ч без отказа. Что можно сказать об оставшейся наработке до отказа?

По приведенному выше рассуждению можно было бы ожидать, что остаточная наработка

$$T^* = T - 20 = 35 \text{ ч}.$$

Однако, что на самом деле показало проведенное испытание наугад выбранного изделия?

Прежде всего то, что это конкретное изделие принадлежит к хорошей партии, характеризующейся наработкой 100 ч. Следовательно, истинное остаточное время работы будет

$$T^* = 100 - 20 = 80 \text{ ч},$$

т. е. оказывается больше первоначального.

В технике для отбраковки «слабых» элементов зачастую испытывают всю партию в ужесточенных по отношению к нормальным режимам, при этом «слабые» элементы «выжигаются», и тем самым оставшиеся работоспособными элементы начинают характеризоваться в среднем лучшими показателями надежности, чем исходная партия элементов. (Конечно, режим таких испытаний, с одной стороны, должен «выжигать» потенциально ненадежные элементы, но с другой, не должен ухудшать характеристик тех элементов, которые успешно проходят такие испытания.)

2.16. Системы из стареющих элементов

Для систем, составленных из «стареющих» элементов, могут быть получены простые и интересные граничные оценки. Предварительно напомним, что в теории надежности последовательной называется система, работоспособность которой обеспечивается только в случае работоспособности всех входящих в ее состав элементов. Если вероятность безотказной работы i-го элемента системы есть $P_i(t)$, то вероятность отказа системы, состоящей из n независимых элементов,

$$P(t) = \prod_{1 \leqslant i \leqslant n} P_i(t).$$

Средняя наработка до отказа такой системы определяется на основании общей формулы

$$T=\int_0^\infty P(t)\,dt=\int_0^\infty \prod_{1\leqslant i\leqslant n} P_i(t)\,dt.$$

Для последовательной системы можно записать следующую очевидную нижнюю границу:

$$P(t)=\begin{cases}\exp\left(-t\sum\limits_{1\leqslant i\leqslant n} T_i^{-1}\right) & \text{для} \quad t\leqslant T^*,\\ 0 & \text{для} \quad t>T^*,\end{cases}$$

где T^* равно наименьшему из средних наработок до отказа элементов системы, т. е.

$$T^*=\min_{1\leqslant i\leqslant n} T_i.$$

Эта нижняя оценка естественным образом следует из (2.26).

Верхняя граница может быть получена из (2.29) в виде

$$P(t)\leqslant\begin{cases}1 \quad \text{для} \quad t\leqslant T^{(1)},\\ \exp\left(-\omega_t^{(1)}t\right) \quad \text{для} \quad T^{(1)}<t\leqslant T^{(2)},\\ \exp\left[-\left(\omega_t^{(1)}+\omega_t^{(2)}\right)t\right] \quad \text{для} \quad T^{(2)}<t\leqslant T^{(3)},\\ \dots\\ \exp\left(-\sum\limits_{1\leqslant i\leqslant n}\omega_t^{(i)}t\right) \quad \text{для} \quad t\geqslant T^{(n)},\end{cases}$$

где $T^{(k)}$ — k-е в порядке возрастания значение T_i, $i=1, \ldots, n$, т. е.

$$T^{(1)}=\min_{1\leqslant i\leqslant n} T_i, \quad T^{(n)}=\max_{1\leqslant i\leqslant n} T_i,$$

а

$$T^{(k)}=\min_{1\leqslant i\leqslant n}\{T_i : T_i\geqslant T^{(k-1)}\}.$$

Показатели экспонент $\omega_t^{(k)}$ связаны с $T^{(k)}$ в соответствии с формулой (2.29).

Параллельной называется система, работоспособность которой обеспечивается работоспособностью одного любого из элементов. Рассмотрим систему, состоящую из одного основного и $n-1$ нагруженных резервных элементов. Верхняя граница для вероятности отказа такой системы

$$Q(t)\leqslant\begin{cases}\prod\limits_{1\leqslant i\leqslant n}\left(1-e^{-t/T_i}\right) & \text{для} \quad t\leqslant T^*,\\ 1 & \text{для} \quad t>T^*,\end{cases}$$

где T^* имеет тот же смысл, что и выше: $Q(t)=1-P(t)$.

Нижняя граница для вероятности безотказной работы

$$P(t)\geqslant\begin{cases}1-\prod\limits_{1\leqslant i\leqslant n}\left(1-e^{-t/T_i}\right) & \text{для} \quad t\leqslant T^*,\\ 0 & \text{для} \quad t>T^*.\end{cases}$$

Верхняя граница для вероятности безотказной работы получается с помощью формулы (2.29) и имеет вид

$$P(t)\leqslant\begin{cases}1 & \text{для} \quad t\leqslant T^{(n)},\\ 1-\prod\limits_{1\leqslant i\leqslant n}\left(1-e^{-\omega_t^{(k)}t}\right) & \text{для} \quad t>T^{(n)}.\end{cases}$$

Для получения оценок средней наработки до отказа последовательной и параллельной систем предварительно докажем следующую лемму[1].

Лемма. Если $g(x)$ — абсолютно интегрируемая функция на положительной полуоси и

$$g(x)\geqslant 0 \text{ при } x<a \text{ и } g(x)\leqslant 0 \text{ при } x>a,$$

при этом

$$\int_0^\infty g(x)\,dx=0,$$

а $f(x)$ неотрицательная и убывает (возрастает), то

$$\int_0^\infty f(x)\,g(x)\,dx\leqslant(\geqslant)\,0.$$

Доказательства этой леммы легко получить, воспользовавшись рис. 2.15, где представлена монотонно возрастающая функция $f(x)$. (Случай убывания $f(x)$ аналогичен.)

Перейдем к получению оценок для средних наработок до отказа систем из стареющих элементов.

Рассмотрим последовательную систему из n «стареющих» элементов. Для такой системы средняя наработка до отказа

$$T=\int_0^\infty\prod_{1\leqslant i\leqslant n} P_i(t)\,dt.$$

Рассмотрим ту же систему, заменив n-й «стареющий» элемент с произвольным распределением наработки на элемент с вырожденным распределением

$$D_n(t)=\begin{cases}1 & \text{для} \quad 0\leqslant t\leqslant T_n,\\ 0 & \text{для} \quad t>T_n.\end{cases}$$

Рассмотрим разность

$$\Delta_n=\int_0^\infty D_n(t)\prod_{1\leqslant i\leqslant n-1} P_i(t)\,dt-\int_0^\infty\prod_{1\leqslant i\leqslant n} P_i(t)\,dt=$$

$$=\int_0^\infty\left[D_n(t)-P_n(t)\right]\prod_{1\leqslant i\leqslant n-1} P_i(t)\,dt.$$

Нетрудно убедиться, что разность $D_n(t)-P_n(t)$ обладает свойствами функции $g(t)$ из приведенной выше леммы, а $\prod\limits_{1\leqslant i\leqslant n-1} P_i(t)$ —

[1] **Соловьев А. Д., Ушаков И. А.** Некоторые оценки для систем из «стареющих» элементов // Автоматика и вычислительная техника. — 1967. — № 6.

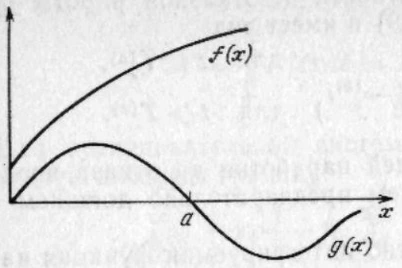

Рис. 2.15. Чертеж к доказательству леммы

монотонная невозрастающая функция. Следовательно, $\Delta_n \geqslant 0$, т. е. замена любого «стареющего» элемента последовательной системы на элемент с вырожденным распределением наработки до отказа (при том же среднем значении наработки) может только повысить среднюю наработку до отказа системы. Последовательно применяя указанную процедуру, получаем верхнюю оценку для средней наработки последовательной системы

$$T \leqslant \min_{1\leqslant i\leqslant n} T_i.$$

Заметим, что эту же оценку можно получить и непосредственно из тех соображений, что для любых $i=1, 2, ..., n$ и $t\geqslant 0$ выполняется условие

$$P_i(t) \geqslant \prod_{1\leqslant i\leqslant n} P_i(t).$$

Нижнюю оценку можно получить, заменив все стареющие элементы в последовательной системе на элементы с экспоненциальным распределением наработки на отказ.

Вновь составим разность

$$\Delta_n = \int_0^\infty \prod_{1\leqslant i\leqslant n} P_i(t)\,dt - \int_0^\infty E_n(t) \prod_{1\leqslant i\leqslant n-1} P_i(t)\,dt = \int_0^\infty [P_n(t)-E_n(t)] \prod_{1\leqslant i\leqslant n-1} P_i(t)\,dt,$$

где $E_n(t)=\exp(-t/T_n)$.

Используя приведенную выше лемму, легко показать, что $\Delta_n\leqslant 0$, т. е. замена любого «стареющего» элемента последовательной системы на элемент с экспоненциальным распределением наработки до отказа при сохранении среднего значения может только уменьшить среднюю наработку до отказа системы. Последовательное применение этой процедуры замены элементов приводит окончательно к условию

$$T=\int_0^\infty \prod_{1\leqslant i\leqslant n} P_i(t)\,dt \geqslant \int_0^\infty \prod_{1\leqslant i\leqslant n} \exp(-t/T_i)\,dt = \left(\sum_{1\leqslant i\leqslant n} T_i^{-1}\right)^{-1}.$$

Для параллельной системы, состоящей из n стареющих элементов, на основании аналогичных рассуждений можно записать следующие нижнюю и верхнюю границы для средней наработки до отказа:

$$\max_{1\leqslant i\leqslant n} T_i \leqslant T \leqslant \int_0^\infty \left\{1-\prod_{1\leqslant i\leqslant n}[1-\exp(-t/T_i)]\right\}dt =$$

$$=\sum_{1\leqslant i\leqslant n} T_i - \sum_{1\leqslant i<j\leqslant n}(T_i^{-1}+T_j^{-1})^{-1}+\ldots(-1)^{n+1}\left(\sum_{1\leqslant i\leqslant n} T_i^{-1}\right)^{-1}.$$

2.17. Статистическая проверка феномена старения

Мы уже отмечали выше, что информация о том, является ли распределение наработки изделия до отказа стареющим или нет, оказывается весьма важной при построении различных практических оценок.

В математической статистике имеется много классических методов оценки принадлежности распределений к тому или иному классу. Для данного конкретного случая можно предложить один интересный специальный прием, позволяющий применить затем для решения нашей задачи известные методы.

Пусть на испытания одновременно поставлено n элементов, которые испытываются до отказа. Испытания прекращаются, когда откажет последний испытываемый элемент. Обозначим: t_k — момент появления k-го по счету отказа в процессе испытаний, а ξ_k — интервалы между моментами появления соседних отказов, т. е.

$$\xi_1=t_1,\ \xi_2=t_2-t_1,\ \ldots,\ \xi_n=t_n-t_{n-1}.$$

Нетрудно представить, что для многих исходных распределений времени работы элементов до отказа величины t_k с малыми индексами будут в среднем меньше тех же величин с большими индексами (если n велико, то это почти очевидно, так как сначала поток отказов вызывается большим числом испытываемых элементов, а по мере испытаний число «источников отказов» убывает из-за появления отказов). Действительно, если бы элементы имели экспоненциальное распределение наработки до отказа, то первый отказ при рассматриваемых испытаниях наблюдался бы в среднем через время $(n\lambda)^{-1}$, второй — в среднем через время $[(n-1)\lambda]^{-1}$, третий — в среднем через время $[(n-2)\lambda]^{-1}$ и т. д. (λ — интенсивность отказов одного элемента.)

Но если случайные величины ξ_k пронормировать определенным образом (каждую из них умножить на число элементов, исправных к началу данного интервала), то полученные нормированные величины $\xi_k^*=(n-k+1)\xi_k$ будут обладать интересными свойствами: если исследуемое распределение наработки элементов до отказа относится к классу «стареющих», то будет выполняться следующее условие для математических ожиданий:

$$M\xi_1^* \geqslant M\xi_2^* \geqslant \ldots \geqslant M\xi_n^*. \qquad (2.31)$$

Однако по результатам испытаний мы получаем все-таки случайные величины ξ_i^*, а не их математические ожидания, поэтому нужно придумать достаточно простые процедуры обработки имен-но этих случайных величин. Сразу же заметим, что если исходное распределение наработки элементов до отказа было экспоненци-альным с параметром λ, то и распределение случайных величин ξ_i^* также будет экспоненциальным с тем же параметром. В этом случае можно одним из известных способов проверить совокуп-ность случайных величин $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ на их принадлежность к классу экспоненциально распределенных. В то же время, так как эти величины в данном случае являются взаимно независимыми одинаково распределенными (по экспоненциальному закону) слу-чайными величинами, можно сказать, что после их упорядочения в порядке возрастания любой порядок первоначальных индексов будет равновероятным.

Если же исходное распределение было «стареющим», то в со-ответствии с условием (2.31) в среднем чаще будут попадаться вначале случайные величины с большими первоначальными но-мерами, а под конец — величины с малыми номерами. Например, понятно, что для сильно «стареющего» распределения «перепуты-вания» номеров в упорядоченной последовательности почти не будет.

Это соображение приводит к возможности использования не-параметрического критерия Вилкоксона для оценки того, относит-ся ли данное исходное распределение к классу «стареющих». Кри-терий Вилкоксона основан на подсчете числа инверсий (инверсией называется ситуация, когда объекты i и $i+1$ занимают места $i+1$ и i). Так, для «стареющего» распределения наиболее вероят-ными являются ситуации, когда число инверсий равно либо нулю (все величины образуют невозрастающую последовательность), либо небольшому числу (среди величин ξ_i^* в среднем наблюдает-ся порядок при незначительных отклонениях). В то же время для экспоненциального распределения число инверсий будет велико, так как любое размещение случайных величин ξ_i^* в упорядочен-ной последовательности является равновероятным.

Соответствующая гипотеза о старении может быть принята с достоверностью 1—β, если число инверсий не превышает заданной границы $\beta n!$ (Максимальное число инверсий равно $n!$) Так, если испытывать, например, пять элементов и при этом хотеть с досто-верностью не ниже 90% оценить, является ли исследуемое рас-пределение стареющим, то из полного числа 120 возможных пе-рестановок лишь 12 с наименьшим числом возможных инверсий будут для нас удовлетворительными; из них одна без инверсий вовсе, пять с одной инверсией и шесть с двумя инверсиями. (Эти шесть должны быть заранее фиксированы, так как всего число различных реализаций с двумя инверсиями в этом случае равно $T_5^2=10$.)

2.18. Оценка надежности ИВС по результатам испытаний ее компонентов

В огромном большинстве случаев при создании сложных сис-тем информатики не удается провести детальных испытаний на надежность, которые позволили бы получить достаточно достовер-ные результаты. Это происходит по многим причинам. К их числу можно отнести прежде всего большие экономические затраты, связанные с проведением таких испытаний, а также недопустимо большое время испытаний, если испытывается достаточно надеж-ная система (при этом в системах с избыточностью могут даже наблюдаться во время испытаний отдельные отказы, которые не приводят к отказу всей системы, а по этой причине не фиксиру-ются).

Кроме того, возможны случаи, когда проведение испытаний системы в целом в некотором фиксированном состоянии принци-пиально невозможно, если сама система находится, например, в постоянном развитии. К таким системам можно отнести большин-ство систем информатики: общегосударственную систему связи, единую систему передачи данных, сеть вычислительных центров, различные информационно-вычислительные системы региональ-ного и отраслевого характера.

Во всех этих случаях приходится оценивать надежность систем по результатам испытаний отдельных ее компонентов. В качестве простейшего примера рассмотрим систему, состоящую из n после-довательно соединенных элементов. Вероятность безотказной ра-боты такой системы за некоторое фиксированное время

$$P=\prod_{1\leqslant i\leqslant n} P_i,$$

где P_i — соответствующая вероятность для i-го элемента системы.

Рассмотрим сначала такой случай: элементы испытываются отдельно, но испытывается ровно N образцов каждого элемента. Пусть число отказов i-го элемента в таких испытаниях составляет d_i. Как можно было бы оценить надежность системы, составлен-ной из этих элементов?

Можно рассуждать следующим образом. Поставим мысленно следующий эксперимент: допустим, что мы собираем систему из тех элементов, которые подвергались испытаниям. В этом случае можно было бы с вероятностью $\hat{P}_i=1-d_i/N$ в эту систему вклю-чить элемент i-го типа, который не отказал бы в составе системы в течение заданного времени. Таким образом, несмещенная оцен-ка для вероятности безотказной работы последовательной систе-мы из n взаимно независимых элементов

$$\hat{P}=\prod_{1\leqslant i\leqslant n} \hat{P}_i.$$

Это означает, что величина $\hat{d}=N\hat{P}$ представляет собой несме-щенную оценку числа отказов системы при N испытаниях. Поскольку величина $\hat{d}$ не обязательно оказывается (и даже с очень большой вероятностью не оказывается) целым числом, то, используя полученное значение $\hat{d}$ при вычислении доверительных оценок для параметров биномиального распределения, следует $\hat{d}$ округлить до ближайшего целого. Причем при вычислении нижней доверительной границы вероятности безотказной работы системы $\hat{d}$ нужно округлять сверху, а при вычислении верхней границы — снизу.

Рассуждения аналогичны и для случая, когда различные элементы испытываются различное число раз (т. е. на испытания ставятся разное число образцов N_i элементов каждого типа).

При этом несмещенная оценка вероятности безотказной работы системы находится по формуле

$$\hat{P} = \prod_{1 \leqslant i \leqslant n} \left(1 - \frac{d_i}{N_i}\right).$$

Затем вычисляется несмещенная оценка числа отказов системы, когда она в целом испытывается число раз, равное минимальному числу образцов отдельно взятых элементов, поставленных на испытания:

$$\hat{d}^* = N^*(1 - \hat{P}),$$

где $N^* = \min\limits_{1 \leqslant i \leqslant n} N_i$.

Дальнейшая доверительная оценка вероятности безотказной работы последовательной системы из независимых элементов производится как для обычных испытаний по схеме Бернулли при числе испытаний N^* и числе отказов $\hat{d}^*$ (с соответствующими округлениями, как и в описанном выше случае).

Справедливость подобной эвристической процедуры обычно обосновывается построением $\prod\limits_{1 \leqslant i \leqslant n} N_1$ всевозможных систем из неиспытываемых элементов и проверкой этих систем на наличие хотя бы одного элемента, который в результате автономных испытаний оказался отказавшим. В этом случае число $\hat{d}^*$ показывает долю общего числа построенных таким образом систем, которые включали хотя бы по одному отказавшему элементу.

Описанная процедура[1] дает очень хорошие практические результаты, однако она не имеет строгого математического обоснования. Оказывается, что для высоконадежных элементов, когда в результате испытаний не появляется ни одного из отказа, удается на строгом математическом уровне получить очень простую и неожиданную на первый взгляд оценку.

[1] **Ллойд Д. К., Липов М.** Надежность: организация исследования, методы, математический аппарат: Пер. с англ. / Под ред. Н. П. Бусленко. — М.: Сов. радио, 1964.

Итак, пусть при испытаниях n элементов не было зафиксировано ни одного отказа. Обозначим это событие A. Нас интересует нижняя доверительная граница для вероятности такого события с коэффициентом доверия α, т. е.

$$\mathcal{P}\{A\} = \prod_{i=1}^{n} \hat{P}_i \geqslant \alpha. \tag{2.32}$$

Задача состоит в нахождении гарантированной нижней доверительной оценки

$$P_\alpha = \min_{\{\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n\}} \prod_{i=1}^{n} \hat{P}_i^{m_i} \tag{2.33}$$

при условии (2.32).

Найдем среди чисел N_i наименьшее

$$N_k = N^* = \min_{1 \leqslant i \leqslant n} N_i$$

и для вероятности с тем же индексом запишем из (2.33) условие

$$P_k \geqslant \alpha^{\frac{1}{N^*}} \prod_{\substack{i=1 \\ i \neq k}} P_i^{-\frac{N_i}{N^*}}.$$

Подставив это выражение в (2.32) для вероятности безотказной работы последовательной системы, получим[1]

$$P \geqslant \min_{P_i} \alpha^{\frac{1}{N^*}} \prod_{\substack{1 < i < n \\ i \neq k}} P_i^{1 - N_i/N^*}. \tag{2.34}$$

Понятно, что минимум правой части этого неравенства из-за того, что при $i \neq k$ $N_i > N_k$, будет достигаться при $P_i = 1$ для $i \neq k$.

Таким образом, получен интересный вывод: если в результате испытаний элементов системы не было зафиксировано ни одного отказа, то нижняя оценка вероятности безотказной работы системы, состоящей из последовательно соединенных элементов, совпадает с соответствующей оценкой для элемента, объем испытаний которого был наименьшим, хотя на первый взгляд могло показаться, что нижняя доверительная оценка надежности системы ниже, чем полученная оценка. Этот факт можно объяснить и на основании следующих простых рассуждений.

Из испытанных элементов можно составить ровно N^* полностью укомплектованных систем. Представим, что именно эти N^* систем и были поставлены на испытания уже в укомплектованном виде. Понятно, что тогда можно говорить о том, что в результате испытаний N^* образцов систем не наблюдалось ни одного отказа, что и позволяет проводить доверительную оценку исходя из наблюдения $d=0$ отказов при испытаниях N^* образцов систем.

[1] **Мирный Р. А., Соловьев А. Д.** Оценка надежности по результатам испытаний ее компонент // Кибернетику — на службу коммунизму. — Т. 2 / Под ред. А. И. Берга, Н. Г. Бруевича, Б. В. Гнеденко. — М.: Энергия, 1964.

Полное изложение различных методов доверительного оценивания показателей надежности систем с более сложной структурой по результатам испытаний отдельных элементов можно найти в [18].

2.19. Статистическое моделирование функционирования ИВС

На практике очень часто приходится иметь дело с такими математическими моделями реальных сложных систем информатики, аналитическое описание которых крайне громоздко либо вообще практически невозможно: система может быть описана лишь через структурное и функциональное взаимодействие своих элементов и при помощи алгоритмов функционирования этой системы в процессе работы. Например, в сетях передачи данных с пакетной коммутацией могут быть такие сетевые протоколы, проверка или оценка соответствующих показателей которых может быть осуществлена только при помощи статистического моделирования.

В других случаях системы допускают аналитическое описание функционирования, однако оно настолько сложно и неконструктивно, что проще дать имитационное описание этих систем. Наконец, статистическое моделирование может служить и просто вычислительным инструментарием при получении конкретных численных результатов, когда другие численные методы оказываются менее эффективными, например, в смысле трудоемкости машинных вычислений или же в смысле подготовки машинного счета (необходима разработка алгоритмов и программ).

Однако есть такая сфера расчетных задач, где использование метода статистического моделирования является едва ли не единственным способом получения количественных характеристик, — это исследование надежности и эффективности функционирования сложных систем информатики.

В качестве примера рассмотрим сложную сеть связи, информация по которой передается в соответствии с динамическими или адаптивными алгоритмами управления, т. е. направление передачи сообщения из транзитного узла определяется не жесткими априорно заданными правилами, а всем текущим состоянием сети (или, во всяком случае, определенной части сети): длинами очередей на соседних узлах связи, временем занятости каналов, характеристиками самого данного сообщения и т. п.

В этом случае, безусловно, проще осуществить имитацию реального процесса на статистической модели, воспользовавшись имеющимися статистическими данными об интенсивностях связи отдельных абонентов сети, о распределениях длительностей занятости каналов, о показателях надежности каналов связи и узлов коммутации. Затем после определенной продолжительности процесса статистического моделирования, в течение которого фиксируются различные текущие характеристики (длительности задержек сообщений, если рассматривается система с ожиданием, или число потерь, если рассматривается система с отказами), на-

копленные события обрабатываются точно так же, как это осуществлялось бы при реальной эксплуатации данной системы связи. При этом, конечно, нужно принимать в расчет, что масштаб времени при моделировании не совпадает, как правило, с реальным, а также то, что в математической модели допущены определенные упрощения и идеализация по отношению к реальному объекту.

Статистическим моделированием называется численный метод решения математических задач при помощи моделирования структур, процессов и других объектов исследования с использованием случайных величин и случайных последовательностей величин с дальнейшей статистической оценкой различных характеристик получаемой совокупности реализаций [1]. Поясним несколько подробнее некоторые термины, использованные в приведенном определении.

Говоря о моделировании структур и процессов, мы имеем в виду, что исследователь сначала должен дать формализованное описание рассматриваемого объекта. Если для аналитических исследований структуры бывает необходимо записывать формулы в замкнутом виде или какие-либо рекуррентные соотношения, то при статистическом моделировании допустимо задание в виде взаимосвязей элементов системы, таблицы их отношений друг к другу по входу-выходу и т. п. При аналитическом описании процесса функционирования исследуемой системы обычно записываются различные системы интегро-дифференциальных уравнений, описывающие динамику изменения состояний системы во времени. При статистическом моделировании процесс функционирования может быть описан просто в терминах алгоритмов функционирования системы и взаимодействия ее элементов во времени.

Случайные величины и их последовательности должны определенным образом генерироваться, причем характер их «случайности» (закон распределения, зависимость между собой, параметрическая зависимость от времени и другие факторы) представляет собой специальную задачу, имеющую вполне конкретное решение в зависимости от специфики исследуемой задачи.

Процесс статистического моделирования представляет собой процедуру многократного повторения определенных условий и взаимодействий элементов системы, причем в результате каждого такого опыта формируется та или иная конкретная реализация исхода такого статистического испытания. После серии опытов исследователь получает некоторую выборку случайных реализаций, которые затем подвергаются стандартным процедурам ста-

[1] **Бусленко Н. П.** Моделирование сложных систем. — М.: Наука, 1968. Решение задач надежности и эксплуатации на универсальных ЭЦВМ / Б. П. Креденцер, М. М. Ластовченко, С. А. Сенецкий, Н. А. Шишонок; Под ред. Н. А. Шишонка. — М.: Сов. радио, 1967.

Полляк Ю. Г. Вероятностное моделирование на ЭВМ. — М.: Сов. радио, 1971.

Бусленко Н. П., Калашников В. В., Коваленко И. Н. Лекции по теории сложных систем. — М.: Сов. радио, 1973.

тистической обработки. (Как правило, если в результате каждой реализации получается некоторая случайная величина, то оценивается ее среднее значение или доверительные границы с заданным коэффициентом доверия.)

Наиболее естественным представляется использование метода статистического моделирования для задач, которые по своей сути имеют вероятностный характер: задач массового обслуживания, управления запасами. Однако этот метод используется очень часто и как чисто вычислительный.

Рассмотрим для наглядной иллюстрации использования метода статистического моделирования простейшую систему с сетевой структурой — так называемую мостиковую схему (см. рис. 2.9). Если известны коэффициенты готовности отдельных элементов такой системы, то не представляет трудности вычисление коэффициента готовности всей системы при условии независимости отдельных ее элементов и по точной формуле, полученной методом перебора.

Для сложных систем с сетевой структурой, включающей в свой состав десятки и сотни элементов, метод статистического моделирования может оказаться одним из основных вычислительных приемов для получения численных значений [1]. Пусть вероятности работоспособного состояния элементов равны 0,5. Для проведения статистического эксперимента воспользуемся табл. 2.1, в которой приведены равномерно распределенные случайные числа. Числа, расположенные в первом столбце таблицы, будем использовать для статистического эксперимента над первым элементом, числа второго столбца — для эксперимента над вторым элементом, и т. д.

В каждом эксперименте будем считать, что реализовано состояние работоспособности для элемента, если соответствующее случайное число θ, умноженное на 10^{-3}, меньше численного значения K_i — коэффициента готовности этого элемента; в противном случае будем считать, что реализуется состояние отказа данного элемента. Иными словами, для каждого элемента вводится индикаторная функция вида

$$\delta_i(\theta)=\begin{cases}1, & \text{если} \quad \theta\leqslant K_i,\\ 0, & \text{если} \quad \theta> K_i.\end{cases}$$

Понятно, что равномерно распределенные случайные числа с вероятностью K_i оказываются меньше K_i, а с вероятностью $1—K_i$ — больше K_i. Состояние рассматриваемой системы в целом характеризуется вектором $\delta=(\delta_1, \delta_2, \delta_3, \delta_4)$, причем каждой случайной реализации вектора δ может быть поставлена в соответствие 1, если система в целом оказывается при этом состоянии элементов работоспособной, и 0 в противном случае.

Результаты статистического эксперимента приведены в табл. 2.2. Из таблицы видно, что для данного эксперимента из 25 опытов 7 исходов соответствуют состоянию отказа системы, т. е. оценка вероятности безотказной работы составляет 0,72. Расчет по точной формуле дает значение

[1] Бусленко Н. П., Калашников В. В., Коваленко И. Н. Лекции по теории сложных систем. — М.: Сов. радио, 1973.

Таблица 2.1. **Равномерно распределенные случайные числа**

218	167	737	498	237
155	844	335	308	003
512	480	704	201	345
635	342	658	752	481
331	485	402	864	398
205	360	181	402	657
682	674	459	904	295
831	304	043	132	766
705	767	185	517	215
564	532	937	594	319
825	354	514	505	152
613	739	043	284	528
131	485	232	356	453
515	248	325	786	176
424	429	930	479	394
340	716	108	056	839
654	958	790	703	499
805	154	202	368	925
337	376	421	473	701
931	069	322	301	595
217	258	688	308	632
684	352	164	744	611
148	723	383	913	102
634	633	136	102	389
048	758	603	204	857

Таблица 2.2. **Результаты статистического эксперимента**

Номер эксперимента	Состояние элемента					Индикатор системы
	1	2	3	4	5	
1	1	1	0	1	1	1
2	1	0	1	1	1	1
3	0	1	0	1	1	1
4	0	1	0	0	1	1
5	1	1	1	0	1	1
6	1	1	1	1	0	1
7	0	0	1	0	1	0
8	0	1	1	1	0	1
9	0	0	1	0	1	0
10	0	0	0	0	1	0
11	0	1	0	0	1	1
12	0	0	1	1	0	0
13	1	1	1	1	1	1
14	0	1	1	0	1	1
15	1	1	0	1	1	1
16	1	0	1	1	0	1
17	0	0	0	0	1	0
18	0	1	1	1	0	1
19	1	1	1	1	0	1
20	0	1	1	1	0	1
21	1	1	0	1	0	1
22	0	1	1	0	0	0
23	1	0	1	0	1	1
24	0	0	1	1	1	0
25	1	0	0	1	0	1

$$(0{,}5)^5+5(0{,}5)^4 0{,}5+8(0{,}5)^3(0{,}5)^2+$$
$$+2(0{,}5)^2(0{,}5)^3=(0{,}5)^5(1+5+8+2)=0{,}5.$$

Этот пример показывает, насколько существенной может быть погрешность при статистическом моделировании (особенно при незначительном числе экспериментов).

Однако если рассматриваемая сеть имеет очень сложную структуру, а элементы ее достаточно надежны, то статистическое моделирование может продолжаться очень долго до получения первого «значащего» исхода эксперимента — отказа системы. В то же время можно заметить, что некоторые возможные реализации эксперимента можно заведомо и не моделировать. Например, если при моделировании состояния мостиковой схемы получается реализация с одним отказавшим элементом или вовсе без отказов, то можно и не проверять факт связности: в данной структуре минимальный разрез состоит из двух элементов, поэтому отказ одного любого элемента не может привести к отказу системы. (Для моделирования мостиковой схемы это были реализации с номерами 1, 2, 5, 6, 13, 15, 19.)

В то же время реализации 10 и 17 содержат «закритическое» число отказов — по четыре, и в этом случае система находится заведомо в состоянии отказа (число оставшихся работоспособных элементов меньше числа элементов, образующих минимальный простой путь).

Для экономии машинного времени при моделировании сети с целью оценки вероятности связности может быть использован следующий прием. В сети определяется мощность минимального простого разреза. Пусть она равна k, т. е. существуют такие k элементов, что при их отказе отказывает вся сеть, но при отказе любых k—1 элементов сеть сохраняет связность (работоспособность по критерию связности входного и выходного полюсов). Тогда моделирование состояний сети можно начинать с реализаций, когда число отказавших элементов равно k или более. Иными словами, нужно спланировать не прямой имитационный эксперимент, а некую условную схему. Допустим, что мы умеем моделировать состояние сети с отказом в ней ровно m элементов. Для простоты предположим, что все элементы сети идентичны. Тогда оценку вероятности связности сети можно провести с помощью смешанного аналитико-статистического метода по формуле [19]

$$\hat{P} = \sum_{0\leqslant i\leqslant k-1} C_n^i p^{n-i} q^i + \sum_{k\leqslant i\leqslant n} \hat{\Phi}_i C_n^i p^{n-i} q^i, \qquad (2.35)$$

где p — вероятность работоспособного состояния элемента; $q=1-p$; $\hat{\Phi}_i$ — статистическая оценка вероятности связности сети при отказе ровно i элементов (полученная в результате статистического моделирования).

Заметим, что при обычном статистическом моделировании число необходимых экспериментов существенно зависит от надежности элементов, так как при высоконадежных элементах состояния отказа сети могут возникать крайне редко. Предложенный аналитико-статистический метод характеризуется практически постоянной трудоемкостью (хотя есть возможность спланировать эксперимент оптимальным образом, чтобы добиться наибольшей достоверности получаемой оценки).

Проиллюстрируем применение формулы (2.35) на примере с мостиковой схемой. Предварительно опишем схему моделирования отказа элементов. В мостиковой схеме всего 5 элементов. Будем считать, что отказал 1-й элемент при моделировании, если выбрано случайное число, лежащее в диапазоне 000 ... 199; что отказал 2-й элемент, если выбрано случайное число из диапазона 200 ... 399 и т. д. Для генерации случайных чисел используем табл. 2.1.

Сначала будем моделировать отказ ровно двух элементов. Для моделирования используем первый столбец табл. 2.1. Результаты моделирования представим в табл. 2.3, в последней строке которой 1 означает работоспособность, 0 — отказ.

При моделировании разыгрываем два подряд отказавших элемента в системе (если оказывается, что второе случайное число попадает в тот же интервал, что и первое, то из таблицы выбирается следующее число; эти случаи в таблице означены звездочкой).

Таблица 2.3. **Результаты моделирования отказов двух элементов**

Номер эксперимента	1	2	3	4	5	6	7	8	9	10	11
Случайные числа	218 155	512 635	331 682*	831 705	564 825	613 131	515 340*	654 805	337 931	217 684	148 634
Номер отказавшего элемента	(2,1)	(3,4)	(2,4)	(5,4)	(3,5)	(4,1)	(3,2)	(4,5)	(2,5)	(2,4)	(1,4)
Индикатор состояния системы	0	1	1	0	1	1	1	0	1	1	1
Номер эксперимента	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
Случайное число	048 844*	480 342	485 360	674 304	767 532	354 739	485 248	429 716	958 154	376 069	258 723*
Номер отказавшего элемента	(1,5)	(3,2)	(3,2)	(4,2)	(4,3)	(2,4)	(3,2)	(3,4)	(5,1)	(2,1)	(2,4)
Индикатор состояния системы	1	1	1	1	1	1	1	1	1	0	1

Таблица 2.4. **Результаты моделирования отказа трех элементов**

Номер эксперимента	1	2	3	4	5	6	7	8	9
Случайные числа	737 335 402*	181 459 937*	514 043 232	325 930 108	790 202 471	322 688 164	383 136 603	498 308 101	762 864 402
Номер отказавшего элемента	(4,2,3)	(1,3,5)	(3,1,2)	(2,5,1)	(4,2,3)	(2,4,1)	(2,1,4)	(3,2,1)	(4,5,3)
Индикатор состояния системы	0	0	0	0	0	0	0	0	0
Номер эксперимента	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Случайные числа	904 132 517	594 284* 786*	479 056 703	368 473 744*	913 102 481*	398 657 152*	528 176* 394	839 499 701*	595 632 102
Номер отказавшего элемента	(5,1,3)	(3,2,4)	(3,1,4)	(2,3,4)	(5,1,3)	(2,4,1)	(3,1,2)	(5,3,4)	(3,4,1)
Индикатор состояния системы	0	0	1	0	0	0	0	0	1

Из проведенных 22 экспериментов оценка условной вероятности безотказной работы при условии отказа ровно двух элементов составила

$$\hat{\Phi}_2 = 18/22 \approx 0{,}82.$$

(Точное значение $\Phi_2 = 0{,}8$).

Аналогичным образом по таблице случайных чисел, используя столбцы 3—5, разыгрываем реализации отказов ровно трех элементов. Результаты эксперимента сведены в табл. 2.4.

На основании проведенных 18 экспериментов оценка условной вероятности безотказной работы при отказе ровно трех элементов

$$\hat{\Phi}_3 = 2/18 \approx 0{,}11.$$

(Точное значение $\Phi_3 = 0{,}2$.)

Окончательно значение искомой вероятности для мостиковой схемы можно получить *при известных вероятностях работоспособности элементов* p по формуле

$$\hat{P} \approx p^5 + C_5^1 p^4 q + 0{,}82 C_5^2 p^3 q^2 + 0{,}11 C_5^3 p^2 q^3.$$

Точная расчетная формула имеет вид

$$P = p^5 + C_5^1 p^4 q + 0{,}8 C_5^2 p^3 q^2 + 0{,}2 C_5^3 p^2 q^3.$$

Подставив значение $p = 0{,}5$, получим

$$\hat{P} \approx (0{,}5)^5 (1 + 5 + 8{,}2 + 1{,}1) \approx 0{,}48.$$

Однако и при таком способе моделирования возникают «избыточные» реализации: например, в седьмой реализации табл. 2.4 отказ системы возникает уже из-за отказа элементов 1 и 2, поэтому введение третьего отказа элемента 4 является излишним.

В связи с этим можно предложить следующую модификацию изложенного выше метода. Будем вводить отказы в системы нарастающим итогом, проверяя при каждом введенном отказе связность сети. Это позволит остановить каждую реализацию именно *в момент наступления отказа системы*. Тем самым в случае моделирования мостиковой схемы все «неинформативные» состояния при двух отказавших элементах будут «развиты» до состояний отказа системы при очередных отказах элементов, а реализации с тремя отказавшими элементами, в которых уже первых два отказа приводят к отказу системы, будут своевременно остановлены.

В этом случае расчетная формула для двухполюсной сети с минимальным разрезом мощности k примет вид

$$\hat{P} = \sum_{0 \le i \le k-1} C_n^i p^{n-i} q^i + \sum_{k \le i \le n} \frac{\sum_{k \le j \le i} s_j}{N} C_n^i p^{n-i} q^i, \qquad (2.36)$$

где s_j — число реализаций из общего числа N, когда отказ системы наблюдался именно при отказе j-го по счету элемента сети.

Сопоставив формулу (2.36) с формулой (2.35), можно заметить, что

$$\hat{\Phi}_i = \frac{1}{N} \sum_{k \le j \le i} s_j,$$

а обе формулы фактически эквивалентны.

Проиллюстрируем модифицированный метод на той же мостиковой схеме. Будем использовать прежнюю таблицу случайных чисел, читая в ней числа *теперь не по столбцам, а по строкам* (*как и прежде, звездочка означает, что* соответствующее число взято после пропуска неподходящих чисел). Результаты моделирования сведены в табл. 2.5, в последнем столбце которой приведены наборы ребер, отказ которых впервые привел к отказу системы. Индикатор *состояния системы не приводится, поскольку для всех наборов он тождественно* равен 0 в силу построения.

Таблица 2.5. **Результаты моделирования до отказа системы**

Номер эксперимента	Случайные числа	Отказавшие элементы
1	218 167	2 1
2	737 498 237	4 3 2
3	155 844 335	1 5 2
4	308 003	2 1
5	512 704* 201	3 4 2
6	345 635 481	2 4 3
7	331 485 864* 181*	2 3 5 1
8	402 657 904*	3 4 5
9	295 831 043*	2 5 1
10	132 766 517* 215	1 4 3 2
11	564 937* 319 152*	3 5 2 1
12	613 043* 284	4 1 2
13	528 131 232	3 1 2
14	356 453 786*	2 3 4
15	176 424 930*	1 3 5
16	479 394 716*	3 2 4
17	108 839* 654	1 5 4
18	958 790	5 4
19	703 499 805	4 3 5
20	154 202	1 2
21	368 925 471* 701*	2 5 3 4
22	931 069 322	5 1 3
23	301 595 688*	2 3 4
24	308 632 164*	2 4 1
25	744 148* 383*	4 1 2
26	913 102 634	5 1 4
27	633 137 389*	4 1 2

Итак, на основании табл. 2.5 вычисляем

$$\hat{\Phi}_2 = 23/27 \approx 0{,}85$$

(четыре реализации с номерами 1, 4, 18, 20 привели к отказу, а остальные не привели к отказу),

$$\hat{\Phi}_3 = 3/27 \approx 0{,}11$$

(лишь три реализации с номерами 7, 11, 21 не привели к отказу при трех отказах).

Таким образом, для мостиковой схемы

$$\hat{P} \approx p^5 + C_5^1 p^4 q + 0{,}85 C_5^2 p^3 q^2 + 0{,}11 C_5^3 p^2 q^3.$$

При $p=0{,}5$

$$\hat{P} \approx (0{,}5)^5 (1+5+8{,}5+11) \approx 0{,}49.$$

Однако статистическое моделирование является наиболее эффективным при имитации процессов функционирования. Чтобы не затенять сути статистического моделирования, рассмотрим один из простейших случаев: дублированную систему с восстановлением, резервный элемент которой находится в состоянии нагруженного резерва [20]. Пусть в начальный момент оба элемента являются совершенно новыми.

Обозначим через $\xi_i^{(k)}$ *случайную наработку i-го элемента при k-м испытании*, а через $\eta_i^{(k)}$ — случайное время его восстановления при k-м испытании. Для конкретности предположим, что элементы идентичны, т. е. имеют одинаковые распределения *наработок на отказ* $F_1(x)=F_2(x)=F(x)=\mathcal{P}\{\xi \leqslant x\}$ *и одинаковые* распределения времени восстановления $G_1(x)=G_2(x)=G(x)=\mathcal{P}\{\eta \leqslant x\}$.

Процедура статистического моделирования функционирования рассматриваемой дублированной системы может быть описана следующим образом: генерируются случайные величины $\xi_1^{(1)}$ и $\xi_2^{(1)}$, каждая из которых является случайной выборкой из генеральной совокупности величин, имеющих распределение $F(x)$; определяется момент $t^{(1)}$, соответствующий наступлению первого отказа элементов дублированной системы, т. е. $t^{(1)}=\min(\xi_1^{(1)}, \xi_2^{(1)})$. Допустим для конкретности, что $\xi_1^{(1)} > \xi_2^{(1)}$. Далее генерируется случайная величина $\eta_2^{(1)}$, которая является случайной выборкой из совокупности случайных величин с распределением $G(x)$.

Для удобства дальнейшего изложения введем в рассмотрение пары $\{t_i^{(k)}, \delta_i^{(k)}\}$, в которых $t_i^{(k)}$— момент смены верхнего индекса $k-1$ на k для i-го элемента, а $\delta_i^{(k)}$ — индикатор состояния i-го элемента после k-й смены верхнего индекса:

$$\delta_i(t) = \begin{cases} 1, & \text{если в момент } t \ i\text{-й элемент работоспособен или переходит в состояние работоспособности,} \\ 0 & \text{в противном случае (состояние отказа или переход в состояние отказа).} \end{cases}$$

Для дублированной системы удобно ввести индикатор состояния $\delta(t)=(\delta_1(t), \delta_2(t))$.

В *новых обозначениях* $t_1^{(1)}=\xi_1^{(1)}$, $t_2^{(1)}=\xi_2^{(1)}$, $t^{(1)}=\min(t_1^{(1)}, t_2^{(1)})$, а $\delta(t^{(1)})=(1{,}0)$.

После генерации $\eta_2^{(1)}$ формируем $t_2^{(2)}=t_2^{(1)}+\eta_2^{(1)}$. Момент $t_1^{(1)}$ автоматически меняет верхний индекс: $t_1^{(1)} \to t_1^{(2)}$. Далее находим $t^{(2)}=\min(t_1^{(2)}, t_2^{(2)})$. Пусть для определенности $t_1^{(2)} > t_2^{(2)}$ (т. е. $\xi_1^{(1)} > \xi_2^{(1)}+\eta_2^{(1)}$). Тогда $t^{(2)}=t_2^{(2)}$ и $\delta(t^{(2)})=(1{,}1)$.

Указанная процедура продолжается. В массив результатов моделирования заносится момент времени $t^{(k)}$, когда впервые $\delta(t^{(k)})=(0{,}0)$, т. е. когда оба элемента системы оказываются в *состоянии отказа* (*момент начала периода отказа дублированной системы с восстановлением*). Момент времени $t^{(k+1)}$, когда хотя бы один из элементов системы восстанавливается, заносится в массив результатов моделирования как момент конца периода отказа системы и начала очередного периода работоспособности.

Процедура статистического моделирования продолжается до тех пор, пока для системы в целом не будет накоплен достаточный *объем статистической информации, обеспечивающий* достоверность требуемых результирующих оценок.

Проиллюстрируем процедуру статистического моделирования рассмотренной выше дублированной системы с восстановлением конкретным числовым *примером. Пусть распределение наработки на отказ каждого элемента имеет* экспоненциальное распределение со средним, равным 100 ч (случайные реализации соответствуют числам из табл. 2.6), а время восстановления — нормальное *распределение со средним 50 ч и средним квадратическим* 6,7 ч (случайные реализации соответствуют числам из табл. 2.7).

В рассматриваемом эксперименте экспоненциально распределенные случайные числа будем брать, начиная с 6-й строки табл. 2.6, а нормально распределенные случайные числа — начиная с 1-й строки табл. 2.7. Ограничим суммарное время моделирования 550 ч.

Таблица 2.6. **Случайные числа с экспоненциальным распределением**

153	179	031	070	145
187	017	110	118	580
067	073	035	160	107
045	107	042	028	073
111	072	091	015	092
161	103	171	091	042
038	039	078	010	122
018	121	315	202	027
035	026	169	066	156
047	063	006	052	114

Таблица 2.7. **Случайные числа с нормальным распределением**

37,1	32,9	44,8	57,3	49,6
36,1	60,3	41,5	47,7	58,9
47,0	44,1	33,1	41,0	53,1
41,2	72,1	49,0	47,1	44,3
42,1	51,1	45,3	37,8	49,4

Описанная процедура примерительно к данному *случаю иллюстрируется* табл. 2.8, которая не требует специальных пояснений. Из этой таблицы можно определить, что впервые система откажет в момент времени $t^{(6)}=311$ ч, затем выйдет из этого состояния в момент $t^{(7)}=330$ ч и т. д. Иными словами, по табл. 2.8 можно найти следующие результирующие значения длительностей наработок и простоев системы:

$$\xi^{(1)}=311 \text{ ч}, \ \eta^{(1)}=330-311=19 \text{ ч};$$

$$\xi^{(2)}=406-330=76 \text{ ч}, \ \eta^{(2)}=422-406=16 \text{ ч};$$

$$\xi^{(3)}=520-422=98 \text{ ч}, \ \eta^{(3)}=521-520=1 \text{ ч};$$

$$\xi^{(4)}=531-521=10 \text{ ч}, \ \tilde{\eta}^{(4)}=550-531=19 \text{ ч}.$$

Таблица 2.8. **Реализация переходов системы из состояния в состояние**

k	$\xi_1^{(k)}$	$\eta_1^{(k)}$	$\xi_2^{(k)}$	$\eta_2^{(k)}$	$t^{(k)}$	$\delta(t^{(k)})$
0					0	(1,1)
1			103		103	(1,0)
2				37	140	(1,1)
3	161				161	(0,1)
4		33			194	(1,1)
5	91				285	(0,1)
6			171		311	(0,0)
7		45			330	(1,0)
8				57	368	(1,1)
9	42				372	(0,1)
10			38		406	(0,0)
11		50			422	(1,0)
12				36	442	(1,1)
13	39				461	(0,1)
14			78		520	(0,0)
15		60			521	(1,0)
16	10				531	(0,0)
17		(42)		48	568	(0,1)

Здесь последняя величина $\tilde{\eta}^{(4)}$ взята не целиком, поскольку перед началом процесса моделирования было принято решение оборвать процесс при $t=550$ ч.

По результатам проведенного элексперимента оценим следующие показатели системы:

1) средняя наработка

$$\hat{T}=\frac{1}{4}\sum_{1\leq k\leq 4}\xi^{(k)}=\frac{495}{4}\approx 123{,}8 \text{ ч};$$

2) среднее время восстановления

$$\hat{\tau}=\frac{1}{4}\sum_{1<k\leq 4}\eta^{(k)}=\frac{55}{4}\approx 13{,}8 \text{ ч};$$

3) коэффициент готовности

$$\hat{K}=\frac{\sum\limits_{1\leq k\leq 4}\xi^{(k)}}{550}=\frac{495}{550}=0{,}9.$$

Проверим, насколько совпадает значение $\hat{K}$ с истинным. Для независимых элементов

$$K=1-Q^2, \tag{2.37}$$

где Q — коэффициент простоя системы, т. е.

$$Q=\frac{M\eta}{M\eta+M\xi}=\frac{50}{50+100}=\frac{1}{3},$$

откуда

$$K=1-1/9\approx 0{,}89.$$

(Нужно только заметить, что при моделировании мы нашли оценку среднего значения нестационарного коэффициента готовности для условия, когда при $t=0$ оба элемента работоспособны, а при аналитическом расчете по формуле (2.37) вычислили стационарный коэффициент готовности системы.)

3. ОПТИМИЗАЦИОННЫЕ ЗАДАЧИ НАДЕЖНОСТИ ИВС

3.1. Оптимальное резервирование

Как уже отмечалось, резервирование является одним из простых и достаточно эффективных методов повышения надежности. Однако при резервировании возникает задача не только обеспечить заданные показатели надежности, но и добиться этого как можно более экономично, с наименьшими суммарными затратами на резервные элементы для системы в целом либо при заданных ресурсных ограничениях достичь максимально возможной надежности [7, 11, 12 [1]].

В качестве подобных ограничивающих ресурсов можно рассмотреть стоимость, массу, габаритные размеры, потребляемую мощность и т. п. Выбор вида ограничивающего ресурса определяется конкретным типом системы и ее назначением. Обычно удается выделить одну наиболее важную характеристику, которую для краткости назовем «стоимость» вне зависимости от ее физической сущности. Однако на практике встречаются ситуации, когда ограничения накладываются по нескольким ресурсам.

Будем рассматривать системы, представляющие собой последовательное соединение взаимонезависимых участков. Участок системы — это не обязательно конструктивно оформленная автономная часть системы, это такая ее часть, для резервирования которой могут быть использованы однотипные элементы. Например, в задачах обеспечения системы запасными элементами к одному участку системы могут быть отнесены условно все однотипные элементы, если даже они конструктивно разнесены и схемно не связаны между собой. Поэтому иногда удобнее говорить не об участке системы, а о резервной группе.

Рассмотрим сначала случай, когда показатель надежности системы выражается в виде произведения соответствующих показателей надежности отдельных резервных групп:

$$R(x_1, \ldots, x_m)=\prod_{1\leq i\leq m} R_i(x_i),$$

где $R_i(x_i)$ — вероятность безотказной работы (или коэффициент готовности) i-й резервной группы при наличии резерва x_i, или в виде суммы:

$$L(x_1, \ldots, x_m)=\sum_{1<i<m} L_i(x_i),$$

где $L=\log R$, $L_i=\log R_i$.

Кроме того, если надежность системы высока, т. е. $Q_i(x_i)=$

[1] См. также: **Ушаков И. А.** Методы решения простейших задач оптимального резервирования при наличии ограничений. — М.: Сов. радио, 1969.

$=1-R_i(x_i) \ll 1/m$, или, что то же самое, $Q(x_1, ..., x_m) \ll 1$, то можно приближенно записать

$$Q(x_1, ..., x_m) \approx \sum_{1 \leq i \leq m} Q_i(x_i).$$

Обычно в задачах оптимального резервирования предполагается, что стоимость резерва для системы в целом $C(x_1, ..., x_m)$ определяется как

$$C(x_1, ..., x_m) = \sum_{1 \leq i \leq m} C_i(x_i)$$

и, кроме того, сама стоимость резерва i-й резервной группы определяется как

$$C_i(x_i) = c_i x_i,$$

где c_i — стоимость одного элемента i-го типа.

При наличии одного ограничивающего фактора возможны постановки двух следующих задач оптимального резервирования.

1. Прямая задача. Раздельным резервированием системы, состоящей из m резервных групп, добиться того, чтобы показатель надежности был не менее заданного R_0 (или не более заданного показателя Q_0 или L_0) при минимально возможной стоимости резерва в целом, т. е.

$$\min_X \{C(x_1, ..., x_m) \mid R(x_1, ..., x_m) \geqslant R_0\}. \tag{3.1}$$

(Условие $R(X) \geqslant R_0$ можно заменить на $Q(X) \leqslant Q_0$ или $L(X) \leqslant L_0$.)

2. Обратная задача. Раздельным резервированием системы, состоящей из m резервных групп, добиться того, чтобы при максимально возможном показателе надежности системы R (или при минимально возможных показателях Q или L) стоимость всего резерва не превысила заданного значения C_0, т. е.

$$\max_X \{R(x_1, ..., x_m) \mid C(x_1, ..., x_m) \leqslant C_0\}. \tag{3.2}$$

(Функцию $R(X)$ можно заменить на $Q(X)$ или $L(X)$, заменив при этом max на min.) Суть оптимизационной задачи, заключающейся в повышении надежности системы путем резервирования при ограчениях на суммарную стоимость, можно пояснить на двух простых частных случаях. Допустим, что все элементы системы равно надежны и в каждой резервной группе имеется ровно по одному основному элементу. В этом случае приоритет по резервированию сначала получают те группы, элементы которых характеризуются наименьшей стоимостью. Если же элементы имеют равную стоимость, то сначала следует резервировать наименее надежные резервные группы.

В более сложных случаях, когда резервные группы содержат различное число элементов, а сами элементы в различных группах различаются и по показателям надежности, и по стоимости, для определения оптимального состава резервных элементов в системе

требуется использовать специальные алгоритмы решения оптимизационных задач.

Метод наискорейшего покоординатного спуска. Обычно исходные данные задачи не отличаются точностью и достоверностью, поэтому использование строгих методов дискретной оптимизации является с практической точки зрения даже некорректным: строгость метода создает иллюзию точности получаемого результата решения, которой на самом деле нет. Поэтому оправданно применение *приближенных алгоритмов, один из которых — метод наискорейшего покоординатного спуска* [20]. Этот метод является модификацией градиентного метода оптимизации, в которой на каждом шаге вычисляется не градиент, а находятся лишь частные производные, максимальное значение из которых определяет направление движения при очередном шаге.

В общем случае решение вариационных задач градиентным методом заключается в том, что отыскивается значение экстремума некоторой функции путем последовательных шагов из начальной точки по направлению градиента. Удобство этого метода заключается в том, что для решения вариационной задачи нужно знать не аналитическое выражение исследуемой функции, а лишь значения функции и ее первых частных производных в точках, в которые мы попадаем в процессе движения к экстремуму функции.

Представим себе процесс создания оптимальной резервированной системы в виде следующего многошагового процесса. Рассматривается система, состоящая из n подсистем (резервных групп), *причем на начальном этапе процесса предполагается, что* ни у одной из подсистем нет резервных элементов. На первом шаге процесса оптимального построения системы отыскиваем такую подсистему, добавление к которой одного резервного элемента дает наибольший «удельный» выигрыш в приросте показателя надежности системы в целом, т. е. наибольший прирост на единицу стоимости. На втором шаге отыскивается следующая подсистема (включая и ту, к которой только что был добавлен резервный элемент), которая характеризуется тем, что добавление к ней одного резервного элемента дает опять наибольшее относительное приращение результирующего показателя надежности системы в целом. Этой новой подсистемой может быть, вообще говоря, и та же подсистема, что и в первый раз. Аналогичным образом процесс построения оптимальной системы продолжается далее [5].

Допустим, что на некотором N-м шаге построенного таким образом процесса каждая i-я подсистема уже имеет по $x_i^{(N)}$ резервных элементов. Заметим, что если мы на каждом шаге процесса построения оптимальной системы описанным выше способом к соответствующей подсистеме добавляли последовательно по одному элементу, то

$$N = \sum_{i=1}^{n} x_i^{(N)}.$$

(Будем в дальнейшем отмечать различные показатели системы, полученные после проведения N-го шага описанного процесса, верхним индексом N.) Для указанного выше шага N результирующий показатель надежности системы определится как

$$R^{(N)} = R(X^{(N)}) = R(x_1^{(N)}, \dots, x_n^{(N)}) = \prod_{i=1}^{n} R_i(x_i^{(N)}),$$

а суммарная стоимость резервных элементов

$$C^{(N)} = C(X^{(N)}) = C(x_1^{(N)}, \dots, x_n^{(N)}) = \sum_{i=1}^{n} c_i x_i^{(N)}.$$

В соответствии с алгоритмом для выбора направления движения на $(N+1)$-м шаге процесса следует найти

$$\gamma^{(N+1)} = \max \gamma_i^N(x_i^{(N)}) = \max \frac{R(X_i^{(N)}, x_i^{(N)}+1) - R(X^{(N)})}{C(X_i^{(N)}, x_i^{(N)}+1) + C(X^{(N)})},$$

где как уже определялось, X_i есть вектор X без компоненты x_i.

Предварительно выразим в удобной форме:

$$R(X_i^{(N)}, x_i^{(N)}+1) = \frac{R_i(x_i^{(N)}+1)}{R_i(x_i^{(N)})} R(X^{(N)}),$$

$$C(X_i^{(N)}, x_i^{(N)}+1) = c_i + C(X^{(N)}).$$

Отсюда

$$\gamma^{(N+1)} = \max_{1\le i\le n} \gamma_i^{(N)}(x_i^{(N)}) = \max_{1\le i\le n} \frac{1}{c_i}\left[\frac{R_i(x_i^{(N)}+1)}{R_i(x_i^{(N)})} R(X^{(N)}) - \right.$$

$$\left. - R(X_i^{(N)})\right] = R(X^{(N)}) \max_{1\le i\le n} \frac{R_i(x_i^{(N)}+1) - R_i(x_i^{(N)})}{c_i R_i(x_i^{(N)})} =$$

$$= R(X^{(N)}) \max_{1\le i\le n} \gamma_i^{(N)}(x_i^{(N)}),$$

где

$$\gamma_i^{(N)}(x_i^{(N)}) = \frac{R_i(x_i^{(N)}+1) - R_i(x_i^{(N)})}{c_i R_i(x_i^{(N)})}.$$

Поскольку $R(X^{(N)})$ входит сомножителем во все величины $\gamma_i^{(N)}(x_i^{(N)})$ и не влияет на отыскание направления движения, то можно упростить вычислительные процедуры, связанные с проведением процесса построения оптимальной системы. Очевидно, что содержание процесса не изменится, если на $(N+1)$-м шаге процедуры будем двигаться в направлении

$$\gamma^{(N+1)} = \max_{1<i<n} \gamma_i^{(N)}(x_i^{(N)}).$$

Сразу даже трудно представить, насколько подобная модификация алгоритма сокращает и упрощает процедуру процесса. Выигрыш от этой модификации станет совершенно очевидным,

когда мы перейдем к непосредственным вычислениям оптимальных векторов состава системы.

Алгоритм оптимизации может быть описан следующим образом.

1. Вычислить

$$\gamma_i^{(0)}(x_i) = \frac{1}{c_i R_i(x_i^{(0)})}[R_i(x_i^{(0)}+1) - R_i(x_i^{(0)})],$$

причем $x_i^{(0)} = 0$ для $i = 1, 2, \dots, n$.

2. Выбрать наибольшую из величин $\gamma_i^{(0)}$

$$\gamma^{(1)} = \gamma_{k_0}^{(0)}(x_{k_0}^{(0)}) = \max_{1\le i\le n} \gamma_i^{(0)}(x_i^{(0)}).$$

3. Найти

$$x_{k_0}^{(1)} = x_{k_0}^{(0)} + 1.$$

4. Остальные $x_i^{(1)}$ для $i \ne k_0$ получить увеличением на единицу верхнего индекса:

$$x_i^{(1)} = x_i^{(0)} + 0.$$

5. В результате построить новый вектор состава системы

$$X^{(1)} = (X_{k_0}^{(0)}, x_{k_0}^{(1)}).$$

6. Вычислить новое значение

$$\gamma_{k_0}^{(1)}(x_{k_0}^{(1)}) = \frac{1}{c_{k_0} R_{k_0}(x_{k_0}^{(1)})}[R_{k_0}(x_{k_0}^{(1)}+1) - R_{k_0}(x_{k_0}^{(1)})].$$

7. Остальные $\gamma_i^{(1)}(x_i^{(1)})$ для $i \ne k_0$ получить увеличением на единицу верхнего индекса:

$$\gamma_i^{(1)}(x_i^{(1)}) = \gamma_i^{(0)}(x_i^{(0)}).$$

Процесс повторить, начиная с операции 1.

В процессе решения прямой задачи оптимального резервирования необходимо вести контроль значения $R(X^{(k)})$, получающегося на каждом k-м шаге. Процесс решения прекращается на таком шаге N, когда $R(X^{(N-1)}) < R_0 \le R(X^{(N)})$. При этом принимается, что вектор состава системы $X^{(N)}$ является искомым решением.

В процессе решения обратной задачи оптимального резервирования необходимо вести контроль значения $C(X^{(k)})$, получающегося на каждом k-м шаге. Процесс решения прекращается на таком шаге N, когда $C(X^{(N-1)}) \le C_0 < C(X^{(N)})$. При этом принимается, что вектор состава системы $X^{(N-1)}$ является искомым решением.

В ряде практических задач может возникнуть следующая ситуация: $(N-1)$-й шаг еще не дает требуемого решения (например, еще не израсходована вся стоимость или еще не достигнуто требуемое значение вероятности безотказной работы), а N-й шаг оптимизационного процесса требует введения элемента с очень большой стоимостью. Так, при решении первой задачи введение оптимального элемента дает значение показателя надежности системы (например, вероятности безотказной работы), заведомо большее требуемого. В то же время резервирование других (не оптимальных на данном шаге) резервных групп системы обеспечивает требуемое значение вероятности безотказной работы при меньших затратах.

При возникновении подобных ситуаций при решении первой задачи можно рекомендовать следующие методы распределения остатка ресурсов.

1. Если некоторый N-й шаг оптимизационного процесса приводит к существенному перерасходу стоимости (или к достижению излишне высокого значения R), то на N-м шаге исключается из рассмотрения резервная группа, имеющая максимальное значение $\gamma^{(N)}$. Если и в этом случае имеет место существенный перерасход стоимости или достигается излишне высокое значение R, то процедуру следует продолжить в том же направлении, т. е. на N-м шаге исключаются из рассмотрения две группы, имеющие самые большие значения $\gamma^{(N)}$, и т. д.

2. Если некоторый N-й шаг оптимизационного процесса приводит к существенному перерасходу стоимости (или к достижению излишне высокого значения R), то следует «снять» резервный элемент, прибавленный на $(N-1)$-м шаге. Затем после $(N-2)$-го шага продолжить процесс, исключив из рассмотрения одну резервную группу, имеющую максимальное значение $\gamma^{(N)}$. Если и в этом случае имеет место существенный перерасход стоимости или достигается излишне высокое значение R, то процедуру следует продолжить в том же направлении, т. е. после $(N-2)$-го шага исключаются из рассмотрения две группы, имеющие самые большие значения $\gamma^{(N-1)}$, и т. д.

Приближенное решение для высоконадежных систем. Наконец, для большинства практических задач (по крайней мере, для ориентировочных решений) может быть предложен следующий способ, использующий гипотезу о высокой надежности системы, получающейся в результате оптимизации[1]. Пусть требуется при минимально возможных затратах достичь требуемого показателя надежности R_0. Если мы рассматриваем высоконадежную систему, т. е. $Q_0 \ll 1$, то и оптимальное решение удовлетворяет условию $Q_i(x_i^{(N)}) \ll 1$. При этом можно также считать, что справедливо приближенное выражение

$$Q_0 = Q(X^{(N)}) \approx \sum_{i=1}^{n} Q_i(x_i^{(N)}). \tag{3.3}$$

Воспользуемся тем, что на любом достаточно удаленном от начала шага процесса решения, т. е. при достаточно большом расходе резервных элементов, имеет место приближенное равенство всех величин $\gamma_i^{(N)}$ некоторой константе $u^{(N)}$, т. е.

[1] Ушаков И. А. Приближенное решение задач оптимального резервирования // Радиотехника. — 1965. — № 12.

$$\gamma_i^{(N)} = \frac{Q_i(x_i)}{c_i} \approx u^{(N)}. \tag{3.4}$$

Выполнение этого условия следует автоматически из решения оптимизационной задачи и методом наискорейшего покоординатного спуска, и методом неопределенных множителей Лагранжа (заметим, что $u^{(N)}$ фактически и является множителем Лагранжа).

Совместное решение (3.3) и (3.4) дает

$$Q_0 \approx \sum_{i=1}^{n} c_i u^{(N)} = u^{(N)} \sum_{i=1}^{n} c_i, \tag{3.5}$$

откуда

$$u^{(N)} = Q_0 / \sum_{i=1}^{n} c_i.$$

Подставив $u^{(N)}$ в (3.5), получим

$$Q_i(x_i^{(N)}) = \frac{c_i}{\sum_{i=1}^{n} c_i} Q_0.$$

Таким образом, если известна функция $Q_i(x)$, то нетрудно найти

$$x_i^{(N)} = Q_i^{-1}\left(Q_0 \frac{c_i}{\sum_{i=1}^{n} c_i}\right),$$

где Q_i^{-1} — обратная функция относительно Q_i.

Обратную задачу можно решить каким-либо итерационным методом, основываясь на решении прямой задачи.

Оптимальный резерв при заданной стоимости системы можно определить при помощи следующей процедуры:

1) находится начальное значение $\gamma^{(1)}$ следующим образом: вычисляются

$$x^{(1)} = (C_0 - C^{(0)}) \left(\sum_{i=1}^{m} c_i\right)^{-1},$$

$$Q^{(1)} = \sum_{i=1}^{m} Q_i(x^{(1)}),$$

$$\gamma^{(1)} = Q^{(1)} \left(\sum_{i=1}^{m} c_i\right)^{-1};$$

2) для найденного значения $\gamma^{(1)}$ определяются $x_1^{(1)}, \ldots, x_m^{(1)}$ из уравнения $Q_i(x_i) = \gamma^{(1)}$;

3) производится контроль полученной стоимости системы по формуле

$$C^{(1)} = C^{(0)} + \sum_{i=1}^{m} x_i^{(1)} c_i.$$

Если $C^{(1)} > C_0$, то выбирается новое значение $\gamma^{(2)} > \gamma^{(1)}$ и для него, как и раньше, определяются $x_1^{(1)}, \ldots, x_m^{(2)}$. Если же $C^{(1)} < C_0$, то выбирается такое новое значение $\gamma^{(2)}$, для которого $\gamma^{(2)} < \gamma^{(1)}$, и для него аналогичным образом находятся $x_1^{(2)}, \ldots, x_m^{(2)}$.

Для ускорения процесса решения при отыскании очередного значения $C^{(k+1)}$ можно использовать метод линейной интерполяции, выбирая $\gamma^{(k+1)}$ по формуле

$$\gamma^{(k+1)} = \gamma + (\gamma^* - \gamma)(C_0 - C)/(C^* - C),$$

где γ и γ^* — соответственно меньшее и большее значения из $\gamma^{(k)}$ и $\gamma^{(k-1)}$; C и C^* — соответственно меньшее и большее значения из $C^{(k)}$ и $C^{(k-1)}$.

Указанный процесс продолжается до тех пор, пока на N-м шаге процесса не будет найдена величина $C^{(N)}$, практически совпадающая с величиной C_0.

Метод динамического программирования. Для решения различных задач целочисленной оптимизации очень часто применяется метод динамического программирования. Этот же метод с успехом можно использовать и для решения задач оптимального резервирования[1]. Объясняется это в первую очередь тем, что задача оптимального резервирования в предположении независимости отдельных подсистем полностью удовлетворяет требованиям независимости решения на очередном шаге процесса от решения, принятого на предыдущей «траектории» решения, а зависит только от достигнутого состояния и оставшегося ресурса.

Для того чтобы решить прямую задачу оптимального резервирования (3.1), введем функцию Беллмана $\varphi_k(r)$ — оптимальное значение целевого функционала в задаче с k переменными и правой частью ограничения, равной r:

$$\varphi_k(r) = \min\left\{\sum_{1\le i\le k} c_i x_i \,\middle|\, \prod_{i=1}^{n} R_i(x_i) \ge r\right\}.$$

Функция $\varphi_k(r)$ строится рекуррентно при $k=1, \ldots, n$ в соответствии с уравнениями

$$\varphi_1(r) = \min_{y_1 \le x_1 \le z_1} \{c_1 x_1 \mid R_1(x_1) \ge r\}, \quad r \in [R_0, 1], \tag{3.6}$$

$$\varphi_k(R_0) = \min_{y_k < x_k < z_k} \{\varphi_{k-1}(r) + c_k x_k \mid rR_k(x_k) \ge R\}, \quad r \in [R_0, 1]. \tag{3.7}$$

Пусть $x_k^*(R)$, $k=1, \ldots, m$, те значения x_k, на которых достигается минимум в (3.6), (3.7). Очевидно, что $\varphi_m(R_0)$ — это оптимальное

[1] **Беллман Р., Дрейфус С.** Прикладные задачи динамического программирования: Пер. с англ. / Под ред. А. А. Первозванского. — М.: Наука, 1965.

значение целевого функционала задачи (3.1). Решение же этой задачи $x_1^{(0)}, \ldots, x_m^{(0)}$ получается последовательно при $k=m, \ldots, 1$:

$$x_m^0 = x_m^*(R_0), \quad R^{m-1} = R_0/R_m(x_m^0),$$

$$x_k^0 = x_k^*(R^k), \quad R^{k-1} = R^k/R_k(x_k^0).$$

Для решения обратной задачи оптимального резервирования (3.2) вводится новая функция Беллмана

$$\psi_k(C) = \max\left\{\prod_{1\le i\le k} R_i(x_i) \,\middle|\, \sum_{1\le i\le k} c_i x_i \le C\right\},$$

которая рекуррентно по k строится в соответствии с уравнениями

$$\psi_1(C) = \max_{\substack{y_1\le x_1\le z_1,\\ C\in[0,C_0]}} \{R_1(x_1) \mid C_1(x_1) \le C\}, \tag{3.8}$$

$$\psi_k(C) = \max_{\substack{y_k\le x_k\le z_k\\ C\in[0,C_0]}} \{\psi_{k-1}(C) R_k(x_k) \mid c + c_k x_k \le C\}.$$

При построении $\varphi_k(C)$ запоминаются значения $x_k^*(c)$, на которых достигается максимум в (3.8). Решение $x_1^0, \ldots, x_m^0$ задачи (3.2) затем определяется как

$$x_m^0 = x_m^*(C_0); \quad C^{m-1} = C_0 - c_m x_m^0,$$

$$x_k^0 = x_k^*(C^k); \quad \widetilde{C}^{k-1} = C^k - c_k x_k^0, \quad k = m-1, \ldots, 1.$$

При решении задач оптимального резервирования методом динамического программирования в целях сокращения процедур переборного характера очень важно предварительно определить границы возможного изменения оптимального числа резервных элементов. Обозначим эти границы на i-м участке системы через y_i и z_i, т. е. $y_i \le x_i^0 < z_i$.

В прямой задаче оптимального резервирования (3.1) значение нижней границы y_i можно получить непосредственно из условия

$$y_i = \min\{x_i \mid R_i(x_i) \ge R_0\},$$

которое означает, что надежность любой резервной группы системы не может быть ниже требуемого значения надежности всей системы. Верхняя граница для прямой задачи оптимального резервирования может быть легко получена при ограничении на суммарное число резервных элементов исходя из того, что общее резервирование является менее эффективным, чем раздельное (поэлементное):

$$\sum_{1<i<n} z_i = \min\left\{mn \,\middle|\, 1 - \left(1 - \prod_{1\le i\le n} R_i(0)\right)^m \ge R_0\right\}.$$

(Условие записано для нагруженного резерва.)

Нижние границы y_i в обратных задачах оптимального резервирования в общем случае полагаются равными нулю. Тогда верхние границы можно найти из условия

$$z_i \leqslant \left] \frac{C_0}{c_i} \right[,$$

где $\rceil x \lceil$ — целая часть числа x.

Метод универсальных производящих функций. Один из новых способов точного решения задачи оптимального резервирования основан на использовании производящей функции специального вида [21—24]. Предварительно напомним, что производящей функцией дискретного распределения $p_k = \mathcal{P}\{\xi = a_k\}$, где ξ — случайная величина, а a_k — одно из возможных ее значений, называется функция

$$\varphi(W) = \sum_{0<k<\infty} p_k W^{a_k}.$$

В модифицированной производящей функции вместо коэффициента p_k будет использоваться $R(k) = \sum\limits_{0 \leqslant j \leqslant k} p_j$. Модифицированная двумерная производящая функция для i-го участка системы имеет вид

$$\varphi_i(y_i, z) = \sum_{0 \leqslant x_i < \infty} R_i(x_i)\, y_i^{x_i} z^{C_i x_i}.$$

В этой производящей функции, представляющей собой полином от двух переменных y_i и z, коэффициент при s-м по счету слагаемом, равный $R_i(s)$, соответствует значению показателя надежности i-й подсистемы при условии, что имеется ровно s резервных элементов i-го типа. Далее, y_i — есть аргумент, являющийся индикатором числа элементов i-го типа, т. е. показатель при y_i является своеобразным «счетчиком» числа s резервных элементов i-го участка, соответствующих данному коэффициенту $R_i(s)$. Наконец, z есть аргумент модифицированной производящей функции, показатель которого является счетчиком расхода ресурса (например, стоимости) на элементы i-го типа. Модифицированная производящая функция для системы в целом может быть представлена в виде многомерного полинома

$$\varphi(z, y_1, y_2, \dots, y_n) = \prod_{1 \leqslant i \leqslant n} \sum_{0 \leqslant x_i < \infty} R_i(x_i)\, y_i^{x_i} z^{C_i x_i}.$$

Эту функцию можно записать в развернутой форме, представляющей собой сумму слагаемых:

$$\varphi(z, Y) = R(X^{(0)}) + R(X^{(1)})\, Y^{X^{(1)}} z^{(C^{(1)}, X^{(1)})} + + R(X^{(2)})\, Y^{X^{(2)}} z^{(C^{(2)}, X^{(2)})} + \dots . \quad (3.9)$$

Здесь обозначено: $Y = (y_1, \dots, y_n)$; $C^{(s)} = ((C_1^{(s)}, \dots, C_n^{(s)})$; $X^{(s)} = (x_1^{(s)}, \dots, x_n^{(s)})$; $(C^{(s)}, X^{(s)})$ — скалярное произведение вектора-строки $(C_1^{(s)}, \dots, C_n^{(s)})$ на вектор-столбец $(x_1^{(s)}, \dots, x_n^{(s)})$, а также введены условные обозначения

$$Y^{X^{(s)}} = \prod_{1<i<n} y_i^{x_i^{(s)}}, \quad z^{(C^{(s)}, X^{(s)})} = z^{\sum_{1<i<n} C_i^{(s)} x_i^{(s)}}.$$

Каждое слагаемое этого полинома имеет следующий смысл: коэффициент $R(X^{(s)})$ при s-м слагаемом, представляющий собой произведение величин $\prod\limits_{1 \leqslant i \leqslant n} R_i(x_i^{(s)})$, численно равен показателю надежности системы в целом, если в системе резервные элементы будут размещены в соответствии с величинами $x_i^{(s)}$ представляющими собой показатели при соответствующих аргументах y_i, а показатель при z численно равен суммарным затратам ресурса (стоимости) для этого же случая. Рассмотрим сначала процедуру построения так называемой доминирующей последовательности. Перечислим слагаемые полинома, представленного в форме (3.9), в порядке возрастания степеней при z. Из полученного ряда исключим все те элементы, коэффициенты $R(X^{(s)})$ которых не удовлетворяют условию $R(X^{(s)}) > R(X^{(k)})$ при $k < s$. Понятно, что такая процедура позволяет оставить только те элементы, которые соответствуют доминирующей последовательности каждый такой элемент характеризуется обязательным возрастанием показателя надежности при росте суммарных затрат. Степени при y_i, $i = 1, 2, \dots, n$, показывают, каким значениям $x_i^{(s)}$ соответствуют значения показателя надежности и суммарных затрат, удовлетворяющие заданным требованиям.

Таким образом, построена доминирующая последовательность, необходимая для решения прямой и обратной задач оптимального резервирования.

3.2. Обеспечение системы запасными блоками и элементами

Средства вычислительной техники характеризуются высокой степенью унификации. В первую очередь это касается электронных блоков (модулей), которые конструктивно оформляются в виде легкосъемных конструкций — типовых элементов замены (ТЭЗ). Сравнительно небольшое число различных ТЭЗов позволяет обеспечить эффективное во всех смыслах техническое обслуживание вычислительной техники: с одной стороны, малая номенклатура различных ТЭЗов позволяет иметь их в достаточном количестве непосредственно на месте эксплуатации, что обеспечивает высокие показатели надежности, а с другой — тот же фактор позволяет добиваться этого при минимальных экономических затратах.

Чтобы представить, какой выигрыш в необходимом числе запасных ТЭЗов может дать унификация, приведем простой пример. Пусть система включает в свой состав 200 различных съемных модулей по 10 штук каждого типа. Пусть все модули условно одинаковы по показателям надежности и по стоимости. Например, наработка каждого из них составляет 500 ч, а время восстановления 1 ч.

Для обеспечения коэффициента готовности $K = 0{,}99$ системы в целом для каждого типа модуля необходим коэффициент готовности $K_1 = \sqrt[200]{0{,}99} = 0{,}999995$. Таким образом, чтобы реализовать такое значение K_1 нужно 10 работающим модулям придать y_1 резервных (в ненагруженном режиме), где y_1 определяется по формуле [5] $K \cong 1-(10\tau/T)^{y_1+1}$. Отсюда $y_1=3$, а всего в системе должно быть $N=200\,y_1=600$ запасных модулей на систему.

Если же за счет унификации число модулей снижено, допустим, в 5 раз, т. е. в системе будет всего 40 различных типов, каждый из которых повторяется уже по 50 раз, то для $K=0{,}99$ каждого типа понадобится $y_1=5$ запасных модулей. Это означает, что всего на систему нужно будет $N_1=40y_1=200$ запасных модулей. Таким образом, число запасных модулей уменьшилось в 3 раза.

Отказавшие съемные модули поступают на восстановление в ремонтный орган, где осуществляется, если это возможно, замена отказавших в них элементов. Уменьшение номенклатуры съемных модулей, вообще говоря, связано и с уменьшением числа типономиналов комплектующих их элементов. Запасные комплектующие элементы формируют в ремонтном органе периодически пополняемый запас. (Конечно, возможны и экстренные поставки элементов в случае необходимости.)

Проиллюстрируем, насколько выгодна унификация на этом уровне, еще на одном примере.

Пусть в системе используется 300 типономиналов элементов, каждый из которых повторяется в системе по 50 раз. Предположим опять, что по своим характеристикам (надежности и стоимости) элементы идентичны, причем каждый элемент характеризуется экспоненциальным распределением наработки с интенсивностью отказов $\lambda=5\cdot10^{-5}$ отказ/ч. Пополнение запаса элементов в ремонтном органе осуществляется ежемесячно, т. е. при непрерывной работе системы примерно через $t=750$ ч наработки. Среднее число отказов элементов каждого типа составит $\Lambda=50\lambda t=1{,}875$. Для обеспечения достаточности элементов с вероятностью 0,9 за указанный период времени нужное их число x_1 находится по формуле

$$x_1=\left\{x:\min\sum_{k=0}^{x}\frac{\Lambda^k}{k!}\,e^{-\Lambda}\geqslant 0{,}9\right\}.$$

По таблицам распределения Пуассона [19] находим $x_1=5$, а общее число элементов на ремонтной базе должно составить $n=300\,x_1=1500$.

Для уровня обеспеченности 0,99 значения $x_1=7$, $n=300\,x_1=2100$.

Теперь, если за счет унификации число типономиналов элементов сокращено в 5 раз, т. е. до 60 типов, а повторяемость их возросла соответственно в 5 раз, т. е. достигла 250, то $\Lambda=250\lambda t=9{,}375$. Аналогичными расчетами получим для уровня 0,9 значения $x_1=14$ и $n=60\,x_1=840$, а для уровня 0,99 — значения $x_1=18$ и $n=60\,x_1=1080$. Иначе говоря, запас сократился примерно в 1,8 и 2 раз соответственно.

Конечно, приведенные примеры носят чисто иллюстративный характер. В действительности все обстоит гораздо сложнее. Прежде всего, унификация, делая каждый из съемных модулей более универсальным, приводит к тому, что для комплектации системы нужно либо большее число модулей того же объема, либо каждый из модулей должен быть сложнее. Выбор разумной

грани между универсализацией и специализацией — вопрос конкретный, зависящий от многих факторов (массовости выпуска изделий, характера технического обслуживания и т. п.). Однако приведенные выше соображения всегда полезны для инженера-проектировщика.

Мы рассмотрели лишь два частных аспекта проблемы технического обслуживания, причем изолированно. Общая схема (даже весьма упрощенная) гораздо сложнее: в самой системе имеется оперативный нагруженный резерв; резервные ненагруженные ТЭЗы позволяют оперативно осуществлять замену отказавших, но само их восстановление в ремонтном органе может продолжаться сравнительно долго; исчерпание определенных элементов в ремонтном органе может продолжаться сравнительно долго; исчерпание определенных элементов в ремонтном органе ограничивает лишь проведение вполне определенного числа восстановлений модулей; отказавшие модули не всегда могут быть восстановлены в ремонтном органе и возращены в резерв и т. д. Сильно формализованная схема функционирования системы с модульной структурой приведена на рис. 3.1.

Совместный анализ вычислительной системы и системы технического обслуживания аналитическими методами практически невозможен. В то же время метод статистических испытаний, позволяющих в принципе учесть многие факторы их взаимосвязи, весьма удобен для получения численных значений различных показателей системы технического обслуживания, но практически из-за огромной трудоемкости решений не дает возможности построения оптимизационных моделей.

Одним из распространенных способов анализа качества и эффективности системы технического обслуживания является выде-

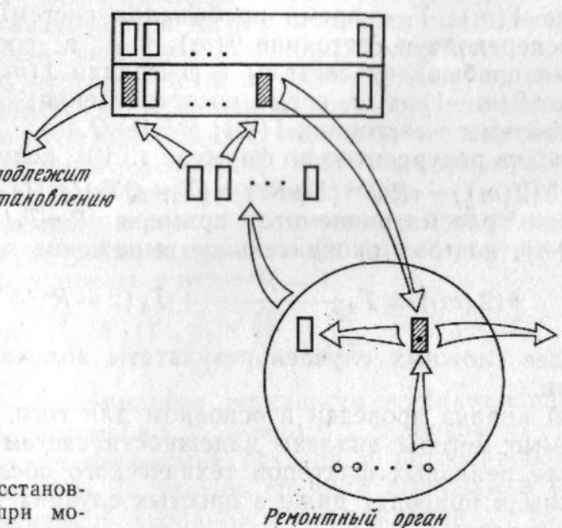

Рис. 3.1. Схема восстановления и ремонта при модульной структуре системы

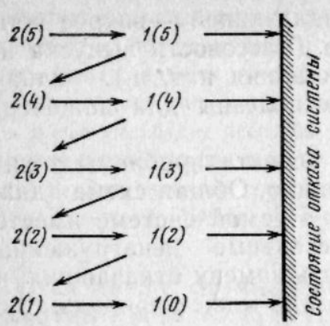

Рис. 3.2. Граф переходов для системы из двух резервных блоков и пяти запасных элементов (в скобках проставлено число имеющихся запасных элементов в данном состоянии)

ление ее в изолированную подсистему, характеризуемую своими собственными показателями. Заметим, что и в этом случае система технического обслуживания остается очень сложным объектом. Рассмотрим сильно идеализированную простейшую частную модель: система состоит из одного рабочего (основного) и одного нагруженного резервного блоков, каждый из которых состоит из идентичных элементов. Имеется запас из 5 элементов, т. е. в системе допускается всего 5 восстановлений, после чего блоки начинают «вымирать». Граф переходов функционирования такой системы представлен на рис. 3.2.

Понятно, что аналитическое решение для определения вероятности безотказной работы даже для такой простой математической модели очень громоздко. Для иллюстрации рассмотрим нахождение средней наработки этой системы. Обозначим $\theta(n(m))$ среднюю наработку системы из состояния $n(m)$, где n — число работоспособных блоков, а m — число оставшихся элементов. Тогда по формуле полного математического ожидания можно записать непосредственно по графу переходов

$$\theta(2(m)) = T_2 + QT_1 + R[\tau + \theta(2(m-1))], \tag{3.10}$$

причем $\theta(2(0)) = T_2 + T_1$. Здесь введены следующие обозначения: T_2 — время пребывания системы в состоянии $2(m)$ до перехода в состояние $1(m)$; T_1 — время пребывания системы в состоянии $1(m)$ до перехода в состояние $\theta(m)$, т. е. в состояние отказа; τ — время пребывания системы в состоянии $1(m)$ до перехода в состояние $2(m-1)$, т. е. время восстановления; Q — вероятность отказа системы в состоянии $1(m)$; $R = 1 - Q$.

Вычислив рекуррентно по формуле (3.10), получим

$$\theta(2(m)) = \tau R Q^{-1}(1 - R^n) + (T_2 + QT_1) Q^{-1}(1 - R^{n+1}).$$

В условиях рассматриваемого примера $R = T_1/(T_1+\tau)$ и $Q = \tau/(T_1+\tau)$, поэтому окончательное выражение можно упростить:

$$\theta(2(m)) = T_2 \frac{1 - R^{n+1}}{Q} + T_1(2 - R^n - R^{n+1}).$$

Для более сложных случаев результаты получаются очень громоздкими.

Такой анализ проведен в основном для того, чтобы показать, что прямые методы анализа надежности систем с учетом более или менее реальных факторов технического обслуживания крайне сложны и приводят даже в простых случаях к очень громоздким результатам.

Попробуем построить более простую модель, но для очень сложной системы, потеряв, конечно, при этом в точности анализа. Пусть система содержит N типов модулей, причем i-го типа модулей в системе Y_i штук, а в резерве находится y_i штук. Вся система сконструирована с использованием n типов элементов, причем j-го типа элементов в системе X_j штук, а в ремонтном органе находится x_j. Обозначим Y, y, X, x вектора с соответствующими компонентами. Нас интересует коэффициент готовности системы, обеспечиваемый как наличием резервных модулей, так и достаточностью запасных элементов, за счет которых осуществляется восстановление отказавших модулей.

Систему технического обслуживания будем считать нормально функционирующей, если она обеспечивает наличие резервных модулей и на рассматриваемом периоде пополнения запасных элементов не происходит нехватки элементов любого типа для восстановления отказавших модулей. Иными словами, вводим показатель именно для системы технического обслуживания — вероятность достаточности резерва модулей и запаса элементов:

$$\Phi = P(Y, y, X, x).$$

Однако и эта задача очень сложная для решения. Поэтому заменим функционал Φ на существенно более простой:

$$\tilde{\Phi} = P(Y, y \mid \infty) P(X, x, t). \tag{3.11}$$

При этом задача разбивается на две изолированные подзадачи: 1) определение вероятности достаточности резервных модулей без ограничения на число восстановлений; 2) определение достаточности занятых элементов за время t.

К сожалению, не удается установить, является ли $\tilde{\Phi}$ оценкой снизу или сверху для Φ: с одной стороны, рассматривая достаточность резерва модулей, мы не ограничиваем число восстановлений, а с другой — недостача резервных элементов какого-либо любого типа на периоде поставки считается нарушением работы системы технического обслуживания, хотя при очередных отказах модулей именно этот тип элемента может и не понадобиться (да и возможность работы системы за счет «вымирания» резервных модулей также не учитывается в этом случае).

В предположении о независимости отказов в системе и марковском характере случайного процесса функционирования функционал (3.11) легко записать в виде

$$\tilde{\Phi} = \prod_{1 \le i \le N} K_i(Y_i, y_i) \prod_{1 \le j \le n} \sum_{0 \le \nu_j \le x_j} P(X_j, \nu_j, t), \tag{3.12}$$

где $K_i(Y_i, y_i)$ — коэффициент готовности восстанавливаемой последовательной системы из Y_i модулей с резервом y_i при известных суммарных интенсивности отказов и интенсивности восстановления i-го модуля (с учетом времени доставки в ремонтный орган) и при заданной дисциплине обслуживания; $P(X_j, \nu_j, t)$ —

вероятность того, что X_j элементов j-го типа за время t откажут ровно ν_j раз.

Все выражения, входящие в (3.12), легко рассчитываются по стандартным формулам [5]. Более того, запись Φ в виде (3.12) позволяет формулировать и оптимизационные задачи, позволяющие находить оптимальные значения y и x.

3.3. Система обеспечения запасными элементами с иерархической структурой

Задачи оптимального резервирования решаются обычно для одного фиксированного периода времени. Однако на практике системы работают в течение длительного времени, что требует больших затрат средств на обеспечение необходимого числа запасных элементов, если требуется достичь высоких показателей надежности. При этом обеспечение высокого заданного показателя надежности означает, что запас крайне редко исчерпывается, причем, как правило, на конец планируемого периода в среднем остается много неизрасходованных элементов. Если учесть, что хранение запасных элементов в условиях обычной эксплуатации приводит зачастую к их быстрому старению, что ведет к потере или ухудшению основных свойств, то это означает, что на следующий плановый период может потребоваться полное обновление запаса.

Чтобы избежать больших потерь и организовать более рациональное обеспечение запасными элементами, целесообразно создание системы складов с иерархической структурой. Прежде чем составить математическую модель такой иерархической структуры запасов элементов, рассмотрим простейшую ситуацию. Пусть некоторый технический объект, состоящий из n идентичных элементов, должен проработать s периодов, причем в начале каждого периода обеспечивается пополнение индивидуального запаса этого объекта до x запасных элементов. Пополнение осуществляется из некоторого центрального склада [25].

Показатель обеспеченности ЗИПом определяется как вероятность того, что на одном из s периодов длительности δ не произойдет нехватки элементов индивидуального запаса и в то же время центральный склад с числом запасных элементов y сможет обеспечить соответствующее пополнение индивидуального запаса. Вначале потребуем выполнения более жестких условий: пусть центральный склад обеспечивает пополнение индивидуального запаса каждый раз до значения x, в противном случае работа

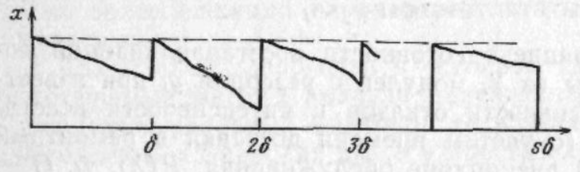

Рис. 3.3. Временнóй график реализации расхода-пополнения элементов индивидуального запаса

центрального склада признается неудовлетворительной (временнóй график расхода-пополнения индивидуального запаса приведен на рис. 3.3).

Если интенсивность отказов каждого элемента равна λ, то вероятность того, что число отказов $\varkappa$ всех n элементов за время δ не превысит x, определяется по формулам распределения Пуассона

$$P(\varkappa \leqslant x) = \sum_{0 \leqslant k \leqslant x} \frac{(n\lambda\delta)^k}{k!} e^{-n\lambda\delta}.$$

Вероятность того, что число отказов равно $\varkappa$ при выполнении условия $\varkappa \leqslant x$, равна

$$\tilde{P}_\varkappa = \frac{(n\lambda\delta)^\varkappa}{\varkappa!} e^{-n\lambda\delta} [P(\varkappa \leqslant x)]^{-1}.$$

Показатель обеспеченности ЗИПом для сформулированных условий

$$P(x, y) = [P(\varkappa \leqslant x)]^s P\left\{\sum_{1 \leqslant k \leqslant s-1} \varkappa^{(k)} \leqslant y\right\}, \qquad (3.13)$$

где $\varkappa^{(k)}$ — число отказов на k-м периоде.

Если вероятность $P(\varkappa \leqslant x)$ велика, то это означает, что условное распределение $\tilde{P}_\varkappa$ практически совпадает с безусловным $P_\varkappa$ распределением Пуассона:

$$P_\varkappa = \frac{(n\lambda\delta)^\varkappa}{\varkappa!} e^{-n\lambda\delta}.$$

Следовательно, можно приближенно записать

$$P\left\{\sum_{1 \leqslant k \leqslant s-1} \varkappa^{(k)} \leqslant y\right\} \approx \sum_{0 \leqslant k \leqslant y} \frac{[n\lambda\delta(s-1)]^k}{k!} e^{-n\lambda\delta(s-1)}.$$

Заметим, что при одном обслуживаемом техническом объекте подобная организация двухступенчатого резерва хуже, чем наличие запаса объема y непосредственно у объекта.

Конечно, формула (3.13) дает слишком жесткую оценку вероятности $P(x, y)$, поскольку даже если

$$\sum_{1 < k \leqslant s-1} \varkappa^{(k)} > y,$$

то это еще не обязательно приводит к нехватке элементов в индивидуальном запасе, так как сам запас x на один период выбирается с достаточным запасом. Формулу (3.13) можно уточнить. В качестве первого приближения можно второй сомножитель в (3.14) представить в виде

$$P(x, y) = [P(\varkappa \leqslant x)]^{s-1} \left[P(\varkappa \leqslant x) P\left\{\sum_{1 \leqslant k \leqslant s-1} \varkappa^{(k)} \leqslant y\right\} + \sum_{1 \leqslant \nu \leqslant x-1} P(\varkappa \leqslant x - \nu) P\left\{\sum_{1 \leqslant k \leqslant s-1} \varkappa^{(k)} = y + \nu\right\}\right], \qquad (3.14)$$

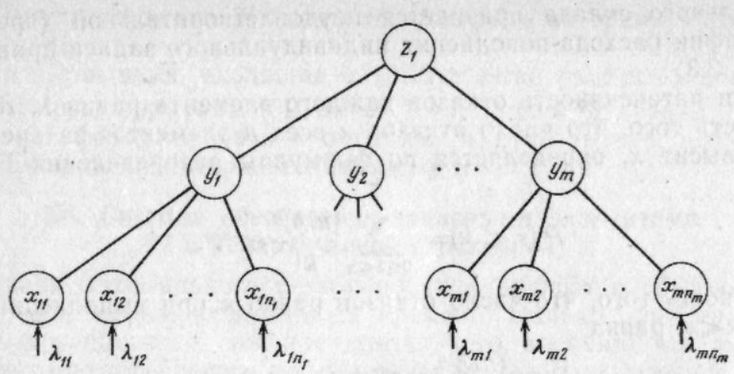

Рис. 3.4. Иерархическая схема запасов идентичных элементов

где ν — число отказов объекта, превысивших уровень запаса центрального склада y.

Формула (3.14) означает, что первые s—1 пополнений осуществляются до уровня x, а на s-м периоде такое пополнение может и не быть осуществлено полностью.

Рассмотрим теперь иерархическую структуру запаса идентичных элементов (рис. 3.4). После выполненных математических заготовок запишем сразу результирующую формулу, используя ради простоты приближенные выражения типа (3.13).

Пусть запас центрального склада равен z и период пополнения равен D, запас j-го регионального склада y_j и период пополнения Δ, запас i-го объекта по j-й ветви регионального снабжения x_{ji} и период пополнения δ. Поток отказов (ji)-го объекта характеризуется интенсивностью λ_{ji}.

Приближенное выражение для вероятности достаточности запасных элементов, т. е. вероятность того, что на периоде D ни на одном уровне иерархии не найдется ни одного запаса, который не удовлетворил бы полностью хотя бы один раз запросы нижнего уровня,

$$P(z, Y, X_1, \ldots, X_m) \approx \prod_{1\leqslant j\leqslant m} \prod_{1\leqslant i\leqslant n_j} \prod_{1\leqslant k\leqslant \frac{D}{\delta}} [P_{ji}(\varkappa_{ji}^{(k)} \leqslant x_{ji})] \times$$

$$\times \prod_{1\leqslant k\leqslant \frac{D}{\delta}} \prod_{1\leqslant j\leqslant m} \left[P_j\left(\sum_{1<k<\frac{\Delta}{\delta}-1} \sum_{1<i<n_i} \varkappa_{ji}^{(k)} \leqslant y_j\right)\right] \times$$

$$\times P\left(\sum_{1\leqslant k\leqslant \frac{D}{\Delta}-1} \sum_{1<j<m} \sum_{1\leqslant l\leqslant n_i} \varkappa_{ji}^{(k)} \leqslant z\right), \qquad (3.15)$$

где $\varkappa_{ji}^{(k)}$ — число отказов (ji)-го объекта на k-м периоде; $X_j = (x_{j1}, \ldots, x_{jn_j})$; $Y = (y_1, \ldots, y_m)$.

Аналогичные схемы могут быть составлены для всех различных типов элементов. В предположении экспоненциальности распределения наработки элементов формула (3.15) может быть существенно упрощена, если при том еще сделать дополнительное предположение о том, что $P(\varkappa_{ji}^{(k)} > x_{ji}) \ll 1$ для любого k:

$$P\left(\sum_{1\leqslant k\leqslant \frac{\Delta}{\delta}-1} \sum_{1\leqslant i\leqslant n_j} \varkappa_{ji}^{(k)} > y_j\right) \ll 1$$

и, наконец,

$$P\left(\sum_{k=1}^{\frac{D}{\Delta}-1} \sum_{j=1}^{m} \sum_{i=1}^{n_j} \varkappa_{ji}^{(k)} > z\right) \ll 1,$$

т. е. можно воспользоваться безусловным распределением числа отказов элементов.

Предположим дополнительно, что все обслуживаемые объекты практически идентичны, т. е. $\lambda_{ji} \approx \lambda$. Тогда формула (3.15) существенно упрощается:

$$P(z, y, x, \lambda) \approx [p_x(\Lambda^{(1)})]^{\frac{D}{\delta} nm} [p_y(\Lambda^{(2)})]^{\frac{D}{\Delta} m} p_z(\Lambda^{(3)}),$$

где n — число объектов, обслуживаемых одним региональным складом;

$$p_x(\Lambda^{(1)}) = \sum_{1\leqslant k\leqslant x} \frac{(\Lambda^{(1)})^k}{k!} e^{-\Lambda^{(1)}}, \quad \Lambda^{(1)} = \lambda\delta;$$

$$p_y(\Lambda^{(2)}) = \sum_{1\leqslant k\leqslant y} \frac{(\Lambda^{(2)})^k}{k!} e^{-\Lambda^{(2)}}. \quad \Lambda^{(2)} = \lambda n(\Delta - \delta);$$

$$p_z(\Lambda^{(3)}) = \sum_{1\leqslant k\leqslant z} \frac{(\Lambda^{(3)})^k}{k!} e^{-\Lambda^{(3)}}, \quad \Lambda^{(3)} = \lambda nm(D - \Delta).$$

Для N типов элементов можно записать вероятность обеспеченности ЗИПами для всей системы технического обслуживания в целом:

$$P(Z, Y, X, \Lambda) = \prod_{1\leqslant l\leqslant N} P(z_l, y_l, x_l, \Lambda_l), \qquad (3.16)$$

где $Z = (z_1, \ldots, z_N)$; z_l — число элементов l-го типа в центральном складе; $Y = (y_1, \ldots, y_l)$; y_l — число элементов l-го типа в региональном складе; $X = (x_1, \ldots, x_l)$; x_l — число элементов на объекте.

Запись формулы (3.16) в развернутой форме (3.15) для случая элементов различного типа не представляет, конечно, никакой сложности, однако она из-за введения еще одного индекса (связанного с типом элемента) становится труднообозримой.

Полученные соотношения могут быть с успехом использованы для решения задач оптимизации системы запасов.

3.4. Оптимальная профилактическая замена элементов на основании информации об их наработке

Как правило, с течением времени у различных технических систем и устройств ухудшаются характеристики функционирования вследствие естественно протекающих процессов старения. При эксплуатации износившиеся устройства и детали обычно заменяют на новые во время специально проводимых профилактических работ. Это связано с тем, что отказ устройства непосредственно во время работы может повлечь за собой определенные дополнительные расходы экономического характера либо привести к опасным последствиям и даже катастрофам.

Подобные предупредительные замены имеют смысл для всех устройств, имеющих стареющее распределение наработки до отказа. Экспоненциальное распределение, являющееся граничным распределением в классе стареющих, не требует таких замен, что совершенно понятно и из общих соображений, так как при экспоненциальном распределении новое устройство ничем по своим вероятностным характеристикам не отличается от уже проработавшего определенное время.

Заметим также, что если замена (или ремонт) отказавшего устройства при аварийном отказе, т. е. непосредственно во время работы, требует тех же затрат, что и во время профилактического ремонта, и не приводит при этом ни к каким дополнительным неприятностям, то, конечно, профилактические замены проводить не имеет никакого смысла.

Если известен вид распределения наработки устройства до отказа, то возможна постановка задачи о назначении периода предупредительных замен из условия минимизации суммарных затрат. Покажем это на следующем простом примере.

Обозначим неизвестный период предупредительных замен через θ. Пусть аварийный отказ устройства приводит к экономическим потерям c_1, а предупредительная замена во время проведения профилактики обходится в c_2 стоимостных единиц. Тогда в течение интервала времени $[0, t]$ средние суммарные затраты

$$C(t, \theta) = c_1 \mathrm{M}\{N_1(t, \theta)\} + c_2 \mathrm{M}\{N_2(t, \theta)\},$$

где $N_1(t_1, \theta)$ — число аварийных отказов устройства за это время; $N_2(t, \theta)$ — число замен, сделанных после безотказной работы устройства в течние периода времени θ.

Если рассматривается достаточно большой период времени $[0, t]$, то можно перейти к рассмотрению задачи о минимизации средних потерь в единицу времени на бесконечном интервале, т. е. минимизировать

$$\widetilde{C}(\theta) = \lim_{t\to\infty} (C(t, \theta)/t).$$

(Кстати, последняя задача существенно проще с точки зрения математического решения, чем первоначальная.)

Заметим, что вместо рассмотрения удельных затрат на бесконечном интервале времени можно рассмотреть средние удельные затраты на одном цикле работы устройства от замены до замены. Можно легко показать, что средние затраты на таком цикле $\widetilde{C}(\theta)$ равны отношению средних затрат на цикле $C(\mathrm{M}\xi)$ к средней длительности такого цикла $\mathrm{M}\xi$.

Средние затраты на цикле

$$C(\theta) = c_1 F(\theta) + c_2 P(\theta),$$

где $P(\theta)$ — вероятность того, что будет обеспечена предупредительная замена (вероятность того, что время работы устройства до отказа окажется больше θ); $F(\theta) = 1 - P(\theta)$ — вероятность того, что замена будет произведена после аварийного отказа (вероятность того, что время работы устройства до отказа окажется меньше θ).

Средняя длительность цикла

$$\mathrm{M}\xi = \int_0^\theta P(t)\,dt.$$

Используя стандартную технику поиска экстремума, находим, что значение θ, минимизирующее средние удельные затраты на бесконечном интервале, определяется из уравнения

$$\frac{d}{d\theta}\left(\frac{c_1 F(\theta) + c_2 P(\theta)}{\int_0^\theta P(t)\,dt}\right) = 0.$$

Его можно переписать в более удобном для данного случая виде:

$$[c_1 f(\theta) - c_2 f(\theta)] \int_0^\theta P(t)\,dt - [c_1 F(\theta) + c_2 P(\theta)]\, P(\theta) = 0.$$

Отсюда интенсивность отказов

$$\lambda(\theta) = \frac{c_1 F(\theta) + c_2 P(\theta)}{(c_1 - c_2)\int_0^\theta P(t)\,dt}$$

или

$$\psi(\theta) = \lambda(\theta) \int_0^\theta P(t) - F(\theta) = \frac{c_2}{c_1 - c_2}.$$

Конечно, могут быть и иные критерии оптимизации периода предупредительных замен. Так, могут быть заданы не стоимости проведения предупредительной и аварийной замены, а их длительности, что приведет к необходимости минимизировать коэффициент простоя устройства (хотя, заметим, математическая постановка в данном случае сохранится с точностью до обозначений) или же можно оптимизировать вероятность выполнения задачи заданной длительности. Могут быть сформулированы и задачи на условную оптимизацию, например добиться заданных эксплуатационных характеристик при минимальных экономических затратах (или добиться максимально возможных эксплуатационных характеристик при заданных экономических затратах).

Постановки задач оптимальных профилактических замен с математической точки зрения существенно *зависят и от того, какая* исходная информация известна исследователю. Например, если допустимы проверки во время эксплуатации, дающие возможность получить некоторые сведения о текущем состоянии контролируемого устройства, то стратегия проведения профилактических мероприятий может быть изменена по сравнению со случаем, когда нет никакой информации, кроме априорной, полученной в момент последней замены [1].

3.5. Оптимальная периодичность контроля работоспособности элементов

Вопрос о проведении проверок работоспособности имеет смысл уже в случае, если контролируемое устройство характеризуется экспоненциальным распределением наработки до отказа [2]. Пусть в результате предварительных испытаний или на основании предыдущей эксплуатации об устройстве нам известно, что распределение наработки между отказами достаточно близко к экспоненциальному с параметром λ.

Представим, что устройство функционирует таким образом, что во время работы мы не можем, не нарушая нормального процесса эксплуатации, удостовериться в том, исправно оно или нет. В то же время нахождение устройства в состоянии необнаруженного отказа — это его простой в самом обычном смысле слова. Мы уже отмечали, что в подобных случаях проведение предупредительных профилактических замен не имеет смысла, так как по своим вероятностным характеристикам новое устройство ничем не лучше того, которое проработало уже в течение произвольного времени. Однако совершенно понятно, что имеет смысл время от времени проводить проверку работоспособности этого устройства, что сокращать возможное время пребывания его в состоянии необнаруженного отказа. Если бы такие проверки не отнимали времени, то их желательно было бы проводить как можно чаще. (Допустим, что сами по себе прерывания работы из-за проверок не играют серьезной роли, а главное — именно потери времени на проведение проверок.) С учетом длительности проведения проверок $t_п$ можно говорить о выборе оптимального периода проверок θ с тем, чтобы $K(\theta, t_п, \lambda)$ — коэффициент использования устройства (доля полезного рабочего времени) был максимальным.

---

[1] **Барзилович Е. Ю., Каштанов В. А.** Некоторые математические вопросы обслуживания сложных систем. — М.: Сов. радио, 1971.

Барзилович Е. Ю., Каштанов В. А. Организация обслуживания при ограниченной информации о надежности системы. — М.: Сов. радио, 1975.

[2] **Ушаков И. А.** Оценка роли аппаратуры контроля в сложных системах // Кибернетику — на службу коммунизму. — Т. 2 / Под ред. А. И. Берга, Н. Г. Бруевича, Б. В. Гнеденко. — М.: Энергия, 1964.

Рассмотрим цикл работы устройства между проверками. С вероятностью $P(\theta) = e^{-\lambda\theta}$ за время θ отказ не произойдет, а с вероятностью $F(\theta) = 1 - e^{-\lambda\theta}$ отказ произойдет до момента проверки, причем распределение случайной наработки ξ в таком интервале будет иметь плотность

$$f_0(t) = \lambda e^{-\lambda t}/(1 - e^{-\lambda\theta}).$$

Среднее значение случайной величины ξ, как нетрудно получить после простых преобразований,

$$M\xi = \theta e^{-\lambda\theta} + \int_0^\theta t\lambda e^{-\lambda t} dt = \frac{1}{\lambda}(1 - e^{-\lambda\theta}).$$

Теперь задача заключается в том, чтобы найти значение θ, обеспечивающее максимум функции

$$K(\theta, t_п, \lambda) = M\xi/(\theta + t_п) = (1 - e^{-\lambda\theta})/\lambda(\theta + t_п).$$

Отыскивая экстремум обычным способом, получаем трансцендентное уравнение

$$e^{\lambda\theta} = 1 + \lambda\theta + \lambda t_п.$$

Итак, зная конкретное значение длительности проверки, можно найти, используя безразмерную величину $\lambda t_п$, значение периода между проверками θ, максимизирующее коэффициент использования устройства.

Заметим, что если $\lambda t_п$ мало (например, $\lambda t_п < 0{,}1$), то с хорошей степенью приближения

$$e^{\lambda\theta} - 1 - \lambda\theta \approx \lambda^2\theta^2/2,$$

откуда

$$Q \approx \sqrt{2t_п/\lambda}.$$

Очень близкая по математическому аппарату задача возникает при организации рестартового режима работы ЭВМ. В обычном режиме любой сбой ЭВМ может привести к неверному окончательному результату, поэтому задачу большого объема разбивают на этапы, в конце каждого из которых осуществляется проверка правильности работы на этом этапе. Если результат правилен, то продолжается решение задачи, а если обнаружены последствия сбоя, то производится повторный счет, но уже только данного этапа. Остановка вычислительного процесса и затраты на проверку правильности снижают производительность ЭВМ, но в то же время предотвращают возможные повторные пересчеты задачи в целом.

Таким образом, возникает задача определения оптимального периода θ проверок правильности текущего счета (т. е. разбиение задачи на этапы), если известны интенсивность сбоев и длительность проверки $t_п$. Понятно, что в такой постановке данная задача полностью эквивалентна ранее рассмотренной.

3.6. Оптимальный поиск отказавших элементов

С ростом сложности систем все острее встает проблема их ремонтопригодности и сокращения времени простоя в процессе эксплуатации. В этой связи одной из важных задач является задача организации оптимальных процедур поиска неисправных (отказавших) элементов, позволяющих быстро локализовать отказ системы.

Конечно, эта задача должна решаться и реально решается с учетом конкретных особенностей той или иной системы, ее структуры, конструктивного оформления, характера эксплуатации, степени обученности эксплуатирующего данную систему технического персонала и других факторов. Однако и в этом случае важно иметь достаточно общие математические модели, описывающие процедуры поиска неисправных элементов в сложных системах, так как такие модели могут помочь сформулировать важные для практики рекомендации [5].

Рассмотрим сначала следующую простую ситуацию[1]: система состоит из n элементов. В данном контексте под элементом будем понимать такую часть системы (обычно конструктивно самостоятельно оформленную), что дальнейшая детализация места отказа будет уже несущественна в рамках данного рассмотрения. Так, при поиске отказа в ЭВМ непосредственно на месте эксплуатации нет необходимости локализовать отказ детальнее, чем до уровня ТЭЗа, который для замены можно взять непосредственно из индивидуального запаса (ЗИП). В то же время в ремонтном органе сам ТЭЗ уже выступает в качестве диагностируемой системы, элементы которой — микромодули, микросхемы, сопротивления, конденсаторы и т. д. — уже подлежат замене, а сами могут и не ремонтироваться дальше.

При проверке системы допускаются тесты двух видов: общий (системный) тест, на основании успешного проведения которого можно сделать заключение об исправности всей системы, и локальный тест каждого отдельного элемента, на основании которого можно сделать заключение об исправности лишь данного проверяемого элемента. В системе возможен отказ нескольких элементов, например, если работоспособность самой системы в целом контролируется не непрерывно, а лишь периодически, причем период этот относительно велик. Требуется, проведя проверки отдельных элементов, выявить все неисправные элементы системы и заменить их в процессе проверки на исправные. (Более сложные тесты нами пока не рассматриваются.)

Задача состоит в том, чтобы на основании имеющейся информации о затратах на проведение общего теста C_0, о затратах на

[1] **Винтер Б.** Оптимальные диагностические процедуры // *Оптимальные задачи надежности*: Пер. с англ. / Под ред. И. А. Ушакова. — М.: Стандарты, 1968.

Беляев Ю. К., Ушаков И. А. Математические модели для задач обнаружения и локализации неисправностей // Кибернетику — на службу коммунизму. — Т. 2 / Под ред. А. И. Берга, Н. Г. Бруевича, Б. В. Гнеденко. — М.: Энергия, 1964.

проведение каждого k-го локального теста, используемого для проверки k-го элемента c_k, вероятности того, что k-й элемент системы является исправным к моменту начала проверки p_k, найти порядок проведения проверок отдельных элементов, обеспечивающий выявление всех имеющихся неисправностей при минимальных в среднем затратах. (Затраты в данном случае могут измеряться в различных единицах — человеко-часах, необходимых для проведения проверок, времени простоя системы, стоимостных единицах и т. п.)

Процедура поиска отказавших элементов будет протекать следующим образом. Все элементы системы нумеруются некоторым образом и в соответствии с этими номерами поочередно проверяются. Как только обнаруживается неисправный элемент, он заменяется на исправный и сразу же проводится общий тест, чтобы удостовериться в том, остались ли в устройстве еще неисправные элементы. Если общий тест не проходит успешно, то продолжается поэлементная проверка.

Сначала для развития интуиции рассмотрим два простых предельных случая. Пусть сначала все элементы равнонадежны, но затраты на их проверку различны. Ясно, что в этом случае проверку следует начинать с элемента, проверка которого требует наименьших затрат. Более того, элементы следует перенумеровать в порядке возрастания затрат на проверку и проверять именно в таком порядке. Теперь предположим, что элементы характеризуются различной надежностью, но все требуют одних и тех же затрат на проверку. В этом случае естественной представляется проверка вначале самого ненадежного элемента. А вся процедура поиска отказавших элементов должна осуществляться в соответствии с номерами элементов, присвоенными последним в порядке возрастания их надежности.

Однако для элементов, различных и по надежности, и по затратам на их проверку, такого однозначного рецепта сразу не возникает. Определим требуемое правило строго, используя для этого так называемый перестановочный прием. Проведем сначала некоторые подготовительные операции.

Затраты, связанные с проверкой некоторого k-го элемента, могут принимать следующие значения: c_k, c_k+C_0 и 0. Первая возможность реализуется в случае, если сам k-й элемент исправен, но подвергается проверке из-за того, что один из элементов с большими номерами неисправен, и поэтому проверка еще не закончилась. Вероятность этого события равна $p_k\left(1-\prod\limits_{k+1\leqslant i\leqslant n} p_i\right)$.

Вторая возможность реализуется, если сам k-й элемент неисправен (тогда он подвергается проверке, а после замены его на исправный обязательно проводится общий тест, чтобы проверить, не был ли этот элемент последним отказавшим). Заметим, что состояние остальных еще не проверенных элементов системы в данном случае несущественно. Вероятность этого события равна $1-p_k$.

Наконец, третья возможность соответствует тому случаю, когда сам k-й элемент и все остальные элементы исправны, т. е. проверка устройства окончилась до проверки k-го элемента. Вероятность этого события не нужна при дальнейших расчетах.

Средние затраты, связанные с проверкой k-го элемента,

$$C_k = c_k p_k \left(1 - \prod_{k+1 \leq i \leq n} p_i\right) + (c_k + C_0)(1 - p_k) =$$

$$= c_k \left(1 - \prod_{k \leq i \leq n} p_i\right) + C_0 (1 - p_k).$$

Таким образом, средние затраты, необходимые для полной проверки устройства, можно вычислить, просуммировав средние затраты по всем элементам:

$$C_\Sigma = \sum_{1 \leq k \leq n} \left[c_k \left(1 - \prod_{k \leq i \leq n} p_i\right) + C_0 (1 - p_k)\right] = \tau - \sum_{1 \leq k \leq n} c_k \prod_{k \leq i \leq n} p_i,$$

где τ — некоторая компонента затрат, не зависящая от нумерации элементов:

$$\tau = \sum_{1 \leq k \leq n} c_k + C_0 \sum_{1 \leq k \leq n} (1 - p_k).$$

Из полученного выражения видно, что средние затраты на проверку всего устройства зависят от того, в каком порядке будем проверять элементы.

Поменяем местами две проверки с соседними номерами, например k и $k+1$, и посмотрим, при каких условиях такая перемена приводит к уменьшению средних суммарных затрат. После перемены местами указанных проверок средние суммарные затраты будут равны:

$$C_\Sigma^* = \tau - \sum_{1 \leq j \leq k+1} c_j \prod_{j \leq i \leq n} p_i - c_{k+1} \prod_{k \leq i \leq n} p_i - \sum_{k+2 \leq i \leq n} c_i \prod_{i \leq j \leq n} p_j.$$

Естественно, что подобная перестановка порядка проверки соседних элементов целесообразна, если выполняется условие

$$C_\Sigma - C_\Sigma^* > 0. \qquad (3.17)$$

Проведя элементарные преобразования, найдем, что условие (3.16) выполняется, если

$$\frac{c_{k+1} p_{k+1}}{1 - p_{k+1}} < \frac{c_k p_k}{1 - p_k}.$$

Ясно, что при нумерации элементов в порядке возрастания величин $c_k p_k (1-p_k)^{-1}$ никакая перестановка не может привести к уменьшению средних суммарных затрат, т. е. такой порядок нумерации является оптимальным в указанном выше смысле.

Интересно заметить, что затраты, связанные непосредственно с устранением отказа, никак не влияют на нумерацию элементов с точки зрения минимизации суммарных средних затрат. Это легко объясняется исходя из того, что затраты на ремонт элемента не зависят от порядка проверки элемента.

На практике возникают и такие ситуации, когда проведение общего теста невозможно по каким-либо причинам. В этом случае возможно проведение локальных тестов уже и для задачи обнаружения отказа за кратчайшее время. Естественно, что сам факт отказа системы предполагается неизвестным. Понятно, что для полностью работоспособной системы в этом случае порядок проверки элементов никакой роли не играет: нужно проверить все элементы. Однако если система содержит хотя бы один отказавший элемент, то встает задача определить этот факт с минимальными потерями. Такая ситуация может возникнуть, если на месте эксплуатации осуществляется только проверка работоспособности системы, а собственно восстановление производится в специально ремонтном органе, где и решается уже задача поиска всех остальных отказавших элементов, если они есть.

Рассмотрим задачу обнаружения отказавших элементов подробнее [1]. Итак, будем считать проведение общего теста невозможным, а факт отказа системы — неизвестным. В этом случае нас интересуют средние суммарные затраты, связанные с проверкой системы при условии, что проверка прекратится именно на k-м элементе, т. е. что первым из обнаруженных отказавших элементов окажется k-й. Искомая величина запишется в виде

$$C_\Sigma = \sum_{1 \leq j \leq n} q_j \prod_{1 \leq i \leq j-1} p_i \sum_{1 \leq i \leq j} c_i.$$

Это выражение можно переписать:

$$C_\Sigma = A_k + q_k \Pi_k (B_k + c_k) + q_{k+1} p_k \Pi_k (B_k + c_k + c_{k+1} + D_k),$$

где

$$A_k = \sum_{1 \leq j \leq k-1} q_i \prod_{1 \leq i \leq j-1} p_i \sum_{1 \leq i \leq j} c_i;$$

$$B_k = \sum_{1 \leq i \leq k-1} c_i;$$

$$D_k = \sum_{k+2 \leq j \leq n} q_i \prod_{1 \leq i \leq j-1} p_i \sum_{1 \leq i \leq j} c_j.$$

Воспользуемся вновь перестановочным приемом: поменяем местами элементы с номерами k и $k+1$ и составим выражение для средних суммарных затрат:

$$C^* = A_k + q_{k+1} \Pi_k (B_k + C_{k+1}) +$$
$$+ q_k p_{k+1} \Pi_k (B_k + c_k + c_{k+1} + D_k).$$

Составим разность $\Delta C = C_\Sigma - C^*$:

[1] **Беляев Ю. К., Ушаков И. А.** Математические модели для решения задач обнаружения и локализации неисправностей // Кибернетику — на службу коммунизму. — Т. 2 / Под ред. А. И. Берга, Н. Г. Бруевича, Б. В. Гнеденко. — М.: Энергия, 1964.

$$\Delta C = \Pi_k [q_k (B_k + c_k) - q_{k+1} (B_k + c_{k+1})] +$$
$$+ \Pi_k [p_k q_{k+1} (B_k + c_k + c_{k+1}) - q_k p_{k+1} (B_k + c_k + c_{k+1})].$$

Неравенство $\Delta C \leqslant 0$ соответствует условию, когда перестановка порядка проверки элементов приводит к ухудшению стоимостных характеристик процедуры.

После элементарных упрощений и приведения подобных членов можно записать это же условие $\Delta C \leqslant 0$ в виде

$$\Delta C = q_{k+1} c_k - q_k c_{k+1} \leqslant 0$$

или

$$c_k / q_k \leqslant c_{k+1} / q_{k+1}.$$

Иначе говоря, нумерация элементов в порядке возрастания величин c_k / q_k обеспечивает минимум средних затрат при поэлементной процедуре обнаружения отказавших элементов в системе.

Задачи обнаружения и поиска отказавших элементов в системе, когда возможно проведение тестов общего вида, можно найти в [26]. Там же решается и задача определения множества достаточных тестов для решения задач технической диагностики.

3.7. Оптимизация на основе статистического моделирования

Метод статистического моделирования, как правило, используется в качестве вычислительной схемы во многих задачах. Известные методы оптимизации с использованием статистических процедур обычно являются процедурами случайного поиска. Грубо говоря, случайный поиск больше всего напоминает метод наискорейшего спуска, при котором направление движения в сторону искомого оптимума осуществляется на основе статистических экспериментов (измерений функции). При этом если от эксперимента к эксперименту те или иные параметры могут принимать случайные значения, то для определения направления движения к оптимуму требуется проведение иногда значительного по объему эксперимента, чтобы выбор был достоверным.

Если исследуется сложная система, то каждая реализация, необходимая для получения численных значений, требует обычно больших затрат машинного времени. Это означает, что процесс движения к оптимуму с использованием статистического поиска может оказаться весьма трудоемким и длительным. Ниже будет продемонстрирован новый прием, который в ряде случаев оказывается более эффективным.

Задача оптимального резервирования. Рассмотрим сначала задачу оптимального резервирования для случая, когда целевой функционал не обладает «удобными» свойствами сепарабельности (разделимости на независимые функции по переменным) и логарифмической выпуклости по каждой переменной, т. е. когда метод динамического программирования в принципе неприменим, а использование метода наискорейшего спуска не обосновано. Таким «неудобным» функционалом является средняя наработка до отказа системы [1].

Итак, пусть имеется последовательная система из n элементов, повышение надежности которой может быть осуществлено за счет применения поэлементного ненагруженного резерва. Стоимость i-го элемента c_i, а распределение наработки до отказа есть $F_i(t) = \mathcal{P}\ \{\xi_i \leqslant t\}$.

Распределение наработки до отказа для i-го участка системы при наличии x_i элементов есть x_i-кратная свертка, т. е.

$$F_i^{*x_i}(t) = P\left\{\sum_{1 \leqslant k \leqslant x_i} \xi_{ik} \leqslant t\right\} = \int_0^t F_i^{*(x_i - 1)}(t-y)\, f_i(y)\, dy, \quad (3.18)$$

где $f_i(y)$ — плотность распределения $F_i(y)$. В рекуррентном уравнении (3.17) по определению $F_i^{*1}(t) = F_i(t)$ и $F_i^{*0}(t) = 1$.

Используя выражение (3.17), можно задачу оптимального резервирования записать в виде

$$\max_X \left\{ \int_0^\infty \prod_{i=1}^n [1 - F_i^{*x_i}(t)]\, dt \;\middle|\; \sum_{1 \leqslant i \leqslant n} c_i x_i \leqslant C_0 \right\},$$

где $X = (x_1, x_2, \ldots, x_n)$.

Таким образом, целевой функционал

$$T(X) = T(x_1, x_2, \ldots, x_n) = \int_0^\infty \prod_{1 \leqslant i \leqslant n} [1 - F^{*x_i}(t)]\, dt$$

нужно максимизировать при ограничениях, наложенных на суммарную кратность сверток. Аналитических методов решения данной задачи не существует.

Применим для решения этой задачи численный метод, существенно использующий процедуру статистических испытаний. Пусть система состоит всего из двух элементов (это условие удобно для простоты графической иллюстрации процесса решения). Стоимости их равны соответственно $c_1 = 1$ и $c_2 = 2$ стоимостных единиц, суммарное ограничение на стоимость системы 8 ст. ед. Распределения времени безотказной работы элементов соответственно $F_1(t)$ — экспоненциальное и $F_2(t)$ — нормальное; реализации случайных величин, представляющих эти распределения, приведены в табл. 3.1, 3.2 (значения их взяты из существующих стандартных таблиц случайных чисел).

Таблица 3.1. Реализации нормально распределенных случайных чисел

37	33	45	57	50
36	60	42	48	59
47	44	33	41	53
41	72	49	47	44
42	51	45	38	49

Таблица 3.2. Реализации экспоненциально распределенных случайных чисел

100	60	20	46	96
125	12	73	75	374
44	49	23	107	72
30	72	28	19	49
74	49	61	10	61

[1] Ушаков И. А., Ясеновец А. В. Статистические методы решения задач оптимального резервирования // Изв. АН СССР. Техн. кибернетика. — 1977. — № 6.

Процедуру статистического моделирования будем проводить следующим образом. В момент $t^{(0)}=0$ генерируются две случайные величины: $\xi_1^{(1)}$ с распределением F_1 и $\xi_2^{(1)}$ с распределением F_2. В момент

$$t^{(1)}=\min(\xi_1^{(1)}, \xi_2^{(1)})$$

осуществляется замена отказавшего элемента. Пусть для конкретности это будет 1-й элемент; тогда в этот момент генерируется новая случайная величина $\xi_1^{(2)}$ с распределением F_1 и находится значение следующего момента отказа 1-го элемента:

$$t_1^{(2)}=\xi_1^{(1)}+\xi_1^{(2)}.$$

В этот же момент величина $t_2^{(1)}=\xi_2^{(1)}$ меняет индекс и автоматически переводится в новое обозначение $t_2^{(2)}$ для удобства описания алгоритма моделирования.

Тогда же в счетчик числа использованных резервных элементов заносится единица в ячейку x_1, т. е. после этого шага в массиве значений x_i будет записано $x_1=1$, $x_2=0$. Далее в массив суммарной стоимости резерва заносится величина c_1.

На втором шаге процесса определяется момент $t^{(2)}=\min(t_1^{(2)}, t_2^{(2)})$ и вся процедура повторяется применительно к тому элементу, который оказался отказавшим в данный момент.

Каждая реализация продолжается до тех пор, пока в счетчике суммарной стоимости резерва не появится величина, превышающая заданное вначале ограничение 8 ст. ед.

Проиллюстрируем описанную процедуру на небольшом примере, записав предварительно нормально и экспоненциально распределенные случайные величины (табл. 3.1, 3.2), используемые в рассматриваемых реализациях статистического моделирования процесса расхода запаса элементов.

Для описания каждой реализации расхода резерва составим отдельные таблицы (табл. 3.3—3.7). В каждой таблице первая строка означает моменты включения элементов 1-го типа, вторая — 2-го типа, на третьей строке в скобках стоит пара накопленных к данному моменту элементов (x_1, x_2), а в четвертой — накопленный расход средств на резервные элементы для системы в

Таблица 3.3. **Первая реализация статистического моделирования расхода элементов**

t_1			100		160		180
t_2	37	70		115		172	
(x_1, x_2)	(0,1)	(0,2)	(1,2)	(1,3)	(2,3)	(2,4)	(3,4)
C	2	4	5	7	8	10	11

Таблица 3.4. **Вторая реализация статистического моделирования расхода элементов**

t_1			125	137			210
t_2	36	96			138	186	
(x_1, x_2)	(0,1)	(0,2)	(1,2)	(2,2)	(2,3)	(2,4)	(3,4)
C	2	4	5	6	8	10	12

Таблица 3.5. **Третья реализация статистического моделирования расхода элементов**

t_1	44		93	116			
t_2		47	91			124	165
(x_1, x_2)	(1,0)	(1,1)	(1,2)	(2,2)	(3,2)	(3,3)	(3,4)
C	1	3	5	6	7	9	11

Таблица 3.6. **Четвертая реализация статистического моделирования расхода элементов**

t_1	30		102		130	149	
t_2		41		113			162
(x_1, x_2)	(1,0)	(1,1)	(2,1)	(2,2)	(3,2)	(4,2)	(4,3)
C	1	3	4	6	7	8	10

целом. (В первой реализации использовались случайные числа, помещенные в первых строках табл. 3.1, 3.2, во второй — числа вторых строк и т. д.)

Для более образного представления реализаций процесса расхода элементов может служить рис. 3.5, на котором представлена также прямая, отсекающая реализации по принятому ограничению 8 ст. ед.

Имея построенные таким образом траектории расхода резервных элементов, можно решать задачу выбора оптимального состава резервных элементов любыми удобными методами целочисленного программирования. В данном случае суть решения можно проиллюстрировать на уровне прямого перебора;

Таблица 3.7. **Пятая реализация статистического моделирования расхода элементов**

t_1		74		123			184
t_2	42		93		138	176	
(x_1, x_2)	(0,1)	(1,1)	(1,2)	(2,2)	(2,3)	(2,4)	(3,4)
C	2	3	5	6	8	10	11

в результате статистического моделирования у нас получалось в примере всего два возможных состава резервных элементов (4,2) и (2,3). Проверим, какое из этих решений является наилучшим для данных реализаций расхода резервных элементов.

Для этого на рис. 3.5 выделим соответствующие прямоугольные площади и определим значения моментов выхода полученных в результате статистического моделирования траекторий из этих областей. Результаты этого анализа, проведенного с помощью и табл. 3.3—3.7, помещены в табл. 3.8.

Из табл. 3.8 следует, что если бы суммарные затраты на резервные элементы были бы распределены как (4,2), то средняя наработка на отказ системы для построенных реализаций статистического моделирования составила бы 125 ч, а при распределении резервных элементов (2,3) — 128,4 ч. Таким образом, второй способ распределения резервных элементов при суммарных затратах 8 ст. ед. более предпочтителен, хотя различие в результатах *и не столь существенно.*

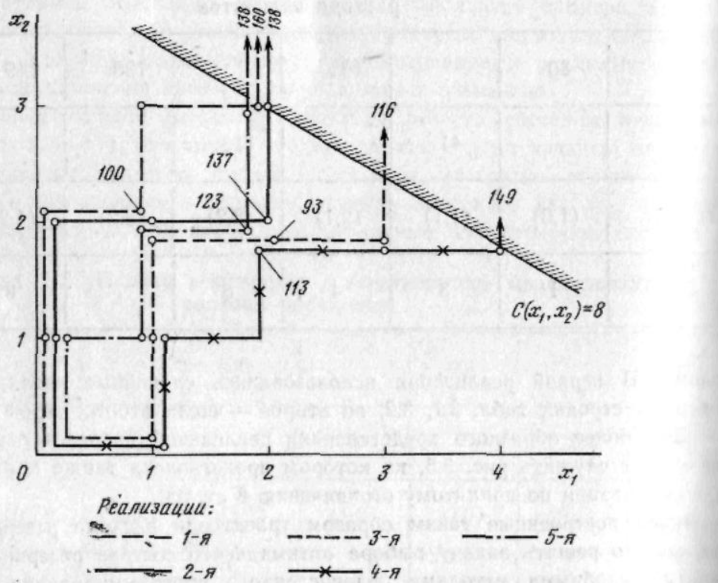

Рис. 3.5. Пример реализации расхода резервных элементов

Таблица 3.8 **Моменты пересечения реализациями статистического моделирования прямоугольных областей с координатами (4,2) и (2.3)**

Область	Моменты для реализации				
	1	2	3	4	5
(4,2)	100	137	116	149	123
(2,3)	160	138	93	113	138

Заметим, что относительно малое число проведенных статистических экспериментов не обеспечивает устойчивости результата, а следовательно, и достоверности решения. Для получения статистически устойчивых решений необходимо провести большее число экспериментов, но при этом начинает существенно возрастать и вычислительная сложность задачи целочисленного программирования.

Оптимизация структуры информационной сети. В качестве второго примера использования статистического моделирования для оптимизации параметров сложной системы приведем решение задачи определения оптимального числа каналов в информационной сети с ненадежными каналами. Рассмотрим систему с поте*рями сообщений: любое сообщение, не нашедшее свободного* работоспособного канала для своей передачи, теряется [31].

Не будем детально описывать процедуру имитационного моделирования процесса функционирования сети связи. В данном контексте даже не важно, какая информация передается (телефонная связь, цифровые данные, пакетные сообщения и т. п.), какого рода коммутация осуществляется (каналов, сообщений, пакетов или комбинированная), какие дисциплины информационного обмена и управления потоками используются (принципы выбора направления передачи, приоритетность, формирование маршрутов и т. п.), наконец, какие функционалы эффективности функционирования выбраны (задержки или потери сообщений или другие критерии качества).

Будем полагать, что имитационная модель с достаточной степенью адекватности соответствует реальной информационной сети. Предположим также, что требуемые значения показателя эффективности функционирования сети можно обеспечить увеличением канальной емкости *(при идентичных по пропускной способности* каналах — увеличением числа каналов на линии связи между соседними узлами).

Задача состоит в том, чтобы добиться заданного значения показателя (для конкретности положим — средней вероятности потерь в сети) при минимальных суммарных затратах (для конкретности положим, что стоимость каналов между любой парой узлов одна и та же). Будем также ради простоты изложения и большей его наглядности считать, что в процессе моделирования развитие сети, о котором речь пойдет ниже, происходит с сохранением исходной структуры: дополнительные каналы могут добавляться только лишь в уже существующие линии связи.

Процедура моделирования осуществляется следующим образом. В качестве исходной берется сеть, у которой каждая линия связи содержит бесконечное число каналов (при практическом осуществлении моделирования можно взять априорно просто достаточно большое число каналов в каждой линии связи или же добавлять их в процессе моделирования по мере необходимости).

В процессе моделирования регистрируется для каждой i-й линии связи $i=1, ..., n$ число занятых в каждый момент времени каналов. Пусть $\nu_i(t)$ — случайное (реализованное в процессе статистического имитационного моделирования) число занятых каналов i-й линии связи в момент времени t. Для каждой i-й линии связи между соседними узлами на основании траектории $\nu_i(t)$ за время моделирования θ может быть составлен ряд величин $k_i^{(m)}$ — числа достижений процессов $\nu_i(t)$ значения $m(\nu_i(t)=m)$, причем именно «снизу вверх».

Пусть за время моделирования θ в сеть передано N сообщений, причем N таково, что $\varepsilon N \gg 1$, где ε — допустимая вероятность потерь сообщений.

Тогда процедуру нахождения оптимального распределения каналов по сети можно описать следующим образом. Пусть в процессе моделирования в течение времени θ зафиксированы следующие максимальные значения числа занимаемых каналов: $(m_1^{(1)}, m_2^{(1)}, ..., m_n^{(1)})$.

Находим минимальную из величин $k_i^{(m_i^{(1)})}$:

$$k^{(1)} = k_j^{(m_j^{(1)})} = \min_{1 \le i \le n} k_i^{(m_i^{(1)})}.$$

Соответствующее значение $m_j^{(1)}$ уменьшаем на 1 и обозначаем $m_j^{(2)}$:

$$m_j^{(2)} = m_j^{(1)} - 1.$$

Остальные $m_i^{(1)}$ для $i \ne j$ переобозначаем: $m_i^{(2)} \to m_i^{(2)}$.

Отыскиваем вновь такой индекс s, для которого

$$k_s^{(2)} = k^{(m_s^{(2)})} = \min_{1 \le i \le n} k_i^{(m_i^{(2)})}.$$

Далее формируется $m_s^{(3)} = m_s^{(2)} - 1$ и т. д.

Процедура продолжается до такого шага M, когда впервые нарушается условие

$$k_\Sigma = \sum_{1 \le l \le M} k^{(l)} < N\varepsilon.$$

Иначе говоря, из сети, которая содержала $(m_1^{(1)}, m_2^{(1)}, ..., m_n^{(1)})$ каналов в соответствующих линиях связи и за время моделирования θ не потеряла ни одной заявки, исключается такое минимально возможное число каналов, которое приведет к потере не более $N\varepsilon$ сообщений из-за недостатка именно числа каналов при выбранной процедуре функционирования сети. Обозначим полученные значения числа каналов в линиях связи после указанной процедуры $\tilde{m}_1^{(1)}, \tilde{m}_2^{(1)}, ..., \tilde{m}_n^{(1)}$. Если период моделирования θ был выбран достаточным, то сформированная сеть по числу каналов будет близка к оптимальной.

Заметим, однако, что пиковые нагрузки на различных линиях связи являются коррелированными величинами: сообщения в сети, образовав в каком-то месте локальный «сгусток», движутся некоторое время в виде своеобразной «волны». Иначе говоря, если за счет исключения из сети канала в какой-либо линии связи мы теряем сообщение при «пиковой» нагрузке, то тем самым, возможно, снижаем и «пиковую» нагрузку на соседних линиях связи. Таким образом, реальное число потерянных заявок может оказаться меньше, чем число k_Σ снятых каналов.

Для уточнения структуры сети с учетом сделанного выше замечания нужно повторить имитационный эксперимент. Удобнее всего в данном случае провести зависимые испытания, а именно повторить ту же реализацию генерации сообщений от различных абонентов сети. (Это легко реализовать, если для генерации моментов формирования сообщений в сети использовать датчик псевдослучайных чисел, который при повторной реализации можно запустить с тех же начальных условий.)

При повторной реализации те же N сообщений пропускаются через сеть с числом каналов $(\tilde{m}_1^{(1)}, \tilde{m}_2^{(1)}, ..., \tilde{m}_n^{(1)})$ и определяется действительное число k_Σ^* потерянных сообщений.

Далее по процедуре, аналогичной описанной выше, из сети вновь изымаются избыточные каналы, но так, чтобы число вновь потерянных сообщений не превысило $k_\Sigma - k_\Sigma^*$.

Подобные итерации могут быть повторены несколько раз. Критерием останова может служить условие совпадения числа потерь сообщений при очередной реализации с числом потерь, получаемым на предыдущей реализации.

Описанный метод позволяет осуществлять оптимизацию параметров сетей без ограничений на тип их структуры и вид ресурсных ограничений. Укажем без детального разбора некоторые дополнительные возможности данного метода. Для обеспечения качества функционирования можно использовать, например, не только дополнительные каналы, но и буферную память (или и то и другое одновременно). При этом может быть построено множество вариантов, оптимальных по Парето (т. е. удовлетворяющих заданным требованиям по показателям качества функционирования при разных соотношениях затрат на каналы связи и на буферную память). Подробнее этот вопрос изложен в [27].

Далее, проектируя сеть по критериям живучести, можно наложить ограничение сверху на допустимое число максимально используемых каналов по тем или иным линиям связи, применяя в остальном ту же процедуру моделирования.

Список литературы

1. **Руденко Ю. Н., Ушаков И. А.** Надежность систем энергетики. — М.: Наука, 1986.
2. **Моделирование** живучести систем энергетики: методология, модель, реализация / М. В. Козлов, Ю. Е. Малашенко, В. С. Рогожин и др. // ВЦ АН СССР. — М., 1986.
3. **Надежность** и живучесть систем связи / Б. Я. Дудник, В. Ф. Овчаренко, В. К. Орлов и др; Под ред. Б. Я. Дудника. — М.: Радио и связь, 1984.
4. **Руденко Ю. Н., Ушаков И. А.** К вопросу оценки живучести сложных систем энергетики // Изв. АН СССР. Энергетика и транспорт. — 1979. — № 1.
5. **Надежность** технических систем: Справочник / Ю. К. Беляев, В. А. Богатырев, В. В. Болотин и др.; Под ред. И. А. Ушакова. — М.: Радио и связь, 1985.
6. **Дзиркал Э. В.** Задание и проверка требований к надежности сложных изделий. — М.: Радио и связь, 1981.
7. **Райкин А. Л.** Элементы теории надежности технических систем / Под ред. И. А. Ушакова. — М.: Сов. радио, 1978.
8. **Соловьев А. Д.** Оценка надежности восстанавливаемых систем. — М.: Знание, 1987.
9. **Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д.** Математические методы в теории надежности. — М.: Наука, 1965.
10. **Ушаков И. А.** Методы расчета эффективности системы на этапе проектирования. — М.: Знание, 1983.
11. **Барлоу Р., Прошан Ф.** Математическая теория надежности: Пер. с англ. / Под ред. Б. В. Гнеденко. — М.: Сов. радио, 1969.
12. **Барлоу Р., Прошан Ф.** Статистическая теория надежности и испытания на безотказность: Пер. с англ. И. А. Ушакова. — М.: Наука, 1984.
13. **Литвак Е. И.** Обобщенное преобразование треугольник-звезда при исследовании свойств сложных сетей // Изв. АН СССР. Техн. кибернетика. — 1981. — № 1.
14. **Ушаков И. А.** Новые оценки характеристик надежности двухполюсных сетей // Надежность и контроль качества. — 1984. — № 5.
15. **Ушаков И. А., Литвак Е. И.** Оценка параметров со сложной структурой // Сообщения при прикладной математике / ВЦ АН СССР. — М., 1986.
16. **Райншке К., Ушаков И. А.** Оценка надежности систем с использованием графов. — М.: Радио и связь, 1988.
17. **Вопросы** математической теории надежности / Е. Ю. Барзилович, Ю. К. Беляев, В. А. Каштанов и др; Под ред. Б. В. Гнеденко. — М.: Радио и связь, 1983.
18. **Павлов И. В.** Статистические методы оценки надежности сложных систем по результатам испытаний. — М.: Радио и связь, 1982.
19. **Козлов Б. А., Ушаков И. А.** Справочник по расчету надежности аппаратуры радиоэлектроники и автоматики. — М.: Сов. радио, 1975.
20. **Горелик В. А., Ушаков И. А.** Исследование операций. — М.: Машиностроение, 1986.
21. **Ушаков И. А.** Универсальная производящая функция // Изв. АН СССР, Техн. кибернетика. — 1986. — № 3.
22. **Ушаков И. А.** Задачи оптимального резервирования и универсальная производящая функция // Изв. АН СССР. Техн. кибернетика. — 1986. — № 6.
23. **Ушаков И. А.** Решение задач многокритериальной дискретной оптимизации с использованием универсальной производящей функции // Изв. АН СССР, Техн. кибернетика. — 1987. — № 1.
24. **Ушаков И. А.** Производящий оператор // Сообщения по прикладной математике / ВЦ АН СССР. — М., 1986.

25. **Ушаков И. А.** Приближенное решение задачи оптимизации иерархической системы многономенклатурного запаса элементов // Надежность и контроль качества. — 1983. — № 6.
26. **Пашковский Г. С.** Задачи оптимального обнаружения и поиска отказов в радиоэлектронной аппаратуре / Под ред. И. А. Ушакова. — М.: Радио и связь, 1981.
27. **Ушаков И. А., Алигулиев Э. А.** Использование метода статистического моделирования для оптимизации числа каналов в сети связи // Изв. АН СССР Техн. кибернетика. — 1988. — № 1.

Содержание