П. А. Зиновьев, Анализ факторов и механизмов живучести в корпоративных информационных системах,
Исслед. по информ., 2007, выпуск 12, 3–30

АНАЛИЗ ФАКТОРОВ И МЕХАНИЗМОВ ЖИВУЧЕСТИ В КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

П.А. Зиновьев

Надежность корпоративных информационных систем (КИС), создаваемых в целях автоматизации производственных и управленческих процессов во многих отраслях, является критически важным фактором, который обязательно необходимо учитывать при их разработке и эксплуатации. Однако даже простая оценка надежности и отказоустойчивости современных крупномасштабных КИС представляет собой весьма серьезную проблему из-за высокой сложности данных объектов [1-3].

В общем плане перед разработчиками КИС стоит не слишком четко сформулированная теоретически и, на первый взгляд, практически неразрешимая задача создания высоконадежных систем, состоящих из малонадежных разнородных (гетерогенных) компонентов. Тем не менее, эти проблемы достаточно успешно преодолеваются за счет применения специальных проектных подходов, нетривиальных расчетных методик и соответствующих архитектурных решений КИС [3-7, 12, 23].

В то же время, в долгосрочной перспективе еще более важной задачей разработчиков КИС может считаться создание сложных корпоративных систем, характеризующихся очень высокой степенью живучести в течение всего жизненного цикла (ЖЦ) этих изделий.

О надежности и живучести сложных технических систем

Рассмотрим стандартную проектную ситуацию: имеется некий набор типовых проектных элементов, выступающих в качестве базовых «строительных блоков» для сложных технических систем (СТС). Для каждого из них известны все основные характеристики, включая показатели надежности (наработка на отказ, время восстановления после сбоев и др.). Необходимо построить из этих блоков систему, которая в перспективе должна будет обладать заведомо более высокими техническими характеристиками, в том числе по надежности, чем входящие в ее состав элементы.

Для построения высоконадежной системы из не обладающих необходимой степенью надежности компонентов самым известным и наиболее типичным конструктивно-технологическим приемом, широко применяемым во многих технических дисциплинах, является избыточность. Избыточность подразумевает резервирование критически важных блоков и уст-

ройств СТС за счет структурного и функционального дублирования, что позволяет поддерживать работоспособность системы при выходе из строя одного или нескольких компонентов. Действительно, параллельное функционирование в составе СТС дублирующих работу друг друга одинаковых блоков и/или устройств обеспечивает, как правило, увеличение степени надежности такой системы по сравнению с надежностью каждого из ее элементов.

В СТС используется как «холодное», так и «горячее» резервирование. «Холодное» резервирование основано на принципе пассивной избыточности, когда резервные блоки и/или устройства включаются в работу только после выхода из строя основных компонентов. «Горячее» резервирование использует принцип активной избыточности, когда резервные элементы функционируют параллельно с основными на постоянной основе, что обеспечивает оперативную взаимозаменяемость, а значит и непрерывность работы критически важных устройств. Примером «горячего» резервирования служит так называемое «зеркалирование» дисковых накопителей, обычно применяемое в корпоративных системах поддержки данных высокой степени готовности [2, 10-12, 18].

Основной недостаток подхода на основе избыточности – значительное увеличение стоимости проектных решений, не гарантирующее при этом пропорциональный рост надежности системы в целом. Тем не менее, для целого ряда особо ответственных СТС такой подход является жизненно необходимым, более того – единственно возможным.

Согласно ГОСТ 34.003-90, надежность (reliability) автоматизированной системы (АС) есть комплексное свойство АС сохранять во времени в установленных пределах значения всех параметров, характеризующих способность АС выполнять свои функции в заданных режимах и условиях эксплуатации [1-3, 11]. Функциональная надежность (dependability) рассматривается как способность АС выполнять заданный набор функций.

К другим важным показателям из этой области, т.е. теории и практики построения надежных СТС, относятся отказоустойчивость, готовность, восстанавливаемость и ремонтопригодность.

Отказоустойчивость можно считать понятием более узким и ограниченным, чем надежность системы в целом, поскольку это качество подразумевает способность системы противостоять определенным классам отказов (процессорные сбои, отказы оперативной и/или дисковой памяти, сбои в работе программного обеспечения и др.) путем целенаправленных мероприятий, ориентированных на борьбу с ними, включая специальные проектные решения.

Готовность (availability) подразумевает гарантированное обеспечение доступности устройств и/или информации в строго определенные интервалы времени в процессе эксплуатации, т.е. предопределяет необходимое со-

отношение между периодами безотказной работы и временем восстановления системы.

Восстанавливаемость и ремонтопригодность (maintainability) характеризуют способность АС восстанавливать свои функции после сбоев и отказов, в том числе путем временного изъятия из эксплуатации отказавших элементов (узлов и блоков) с целью их ремонта и замены резервными элементами. Особый интерес в этом плане представляют системы с возможностями самовосстановления (self-healing).

Так или иначе, для сложных систем с большим количеством компонентов и достаточно сложной конфигурацией бывает очень непросто построить соответствующую математическую модель, позволяющую произвести комплексную оценку надежности СТС, даже если известны все необходимые показатели и параметры для всех ее компонентов. Это особенно актуально для таких сверхсложных СТС, как корпоративные информационные системы.

Следовательно, необходима выработка более перспективных подходов и методов оценки показателей качества функционирования СТС, основанных на комплексных критериях оценки надежностно-стоимостных характеристик. Такие подходы должны позволять более эффективно и адекватно анализировать общее состояние системы с позиций надежности и отказоустойчивости на всем протяжении жизненного цикла данного изделия. Представляется целесообразным в качестве такого наиболее содержательного интегрального показателя использовать «живучесть» СТС.

Под живучестью (survivability) автоматизированной системы (АС), согласно определению, приведенному в ГОСТ 34.003-90, понимается свойство АС, характеризуемое способностью выполнять установленный объем функций в условиях воздействия внешней среды и отказов компонентов системы в заданных пределах [6, 11]. С понятием живучести тесно связано также и такое свойство АС, как адаптивность, определяемое как способность системы изменяться для сохранения своих эксплуатационных показателей в заданных пределах при изменениях внешней среды.

Функциональная живучесть характеризует допустимые пределы снижения качества функционирования КИС, т.е. эффективности выполнения системой заданного набора функций на протяжении всего ЖЦ в условиях постоянной деградации ее ресурсов. Успешное выполнение системой в течение ЖЦ всех предписанных функций и задач в конечном итоге и означает достижение заданных целей ее функционирования. Исследование проблемы обеспечения живучести СТС требует, кроме всего прочего, проведения анализа уязвимостей и управления рисками.

Далее следует разделить проблему обеспечения живучести на структурную и функциональную составляющие. В данной работе рассматриваются в основном функциональные аспекты живучести КИС. Исследование структурных аспектов требует отдельного рассмотрения [27].

Постановка проблемы и описание объекта исследования

Современные корпоративные информационные системы, несомненно, стоят в одном ряду с самыми сложными техническими изделиями, когда-либо созданными человеческими руками и интеллектом. Это крупномасштабные высокотехнологичные территориально-распределенные комплексы, включающие как высокопроизводительные мэйнфреймы и мидфреймы, так и сотни и даже тысячи серверов (точнее, серверных платформ), десятки и сотни тысяч рабочих станций, терабитные базы данных, транснтинентальные телекоммуникационные магистрали, информационные потоки, измеряемые десятками и сотнями гигабит в секунду, колоссальные объемы прикладного программного обеспечения и многое другое.

Для систем со столь сложной архитектурой и необозримым множеством гетерогенных компонентов оценить надежность и живучесть представляется крайне сложной проблемой, даже если известны все необходимые метрики для всех входящих в состав данной корпоративной системы элементов. Более того, в условиях непрерывно и динамично меняющихся требований со стороны внешней бизнес-среды и быстрого морального старения инфокоммуникационных (ИКТ) активов компании подобная задача вообще представляется неразрешимой. Поэтому представляется целесообразным переосмысление подходов к данной проблеме, как ответ на потребность в разработке простых и достаточно эффективных с инженерной точки зрения методик, обеспечивающих адекватно достоверную оценку основных характеристик надежности и живучести современных КИС.

Прежде всего, в составе КИС следует выделить две функционально отличающиеся и, соответственно, принципиально по-разному строящиеся архитектурные составляющие, а именно так называемые «front-end» и «back-end» подсистемы, т.е. подсистемы «переднего» и «заднего» плана. К подсистемам переднего плана относятся архитектурные компоненты КИС, предназначенные для работы с внешними пользователями (Web-серверы, прикладные программные компоненты подсистем CRM – Customer Relation Management и др.). К подсистемам «back-end» относятся компоненты intranet-архитектуры: БД-серверы (СУБД и хранилища данных), серверы и программные средства поддержки корпоративных бизнес-приложений, аналитики, отчетности и т.д., иначе говоря, ИКТ-средства, предназначенные для поддержки внутрикорпоративных механизмов функциональной деятельности.

В рамках тенденции виртуализации гетерогенные ИКТ-активы КИС рассматриваются как информационно-вычислительные среды, приводимые к однородным. Тогда имеет смысл рассматривать живучесть КИС как зависящую от применяемых функционально-логических модельных представлений реальных ИКТ-ресурсов. Будем считать, что КИС обладает следующими категориями виртуальных ИКТ-ресурсов:

- вертикально масштабируемые ресурсы обработки данных (мэйнфреймы, SMP-серверы, серверные кластеры и т.д.);

- горизонтально масштабируемые вычислительные ресурсы («лезвийные» серверы, серверные «фермы», рабочие станции и др.);

- общесистемные и прикладные программные ресурсы (СУБД, OLTP- и OLAP-приложения, ERP-приложения, web-сервисы, прикладные и инфраструктурные сервисы);

- ресурсы хранения данных (традиционные дисковые накопители, сети SAN, дисковые массивы RAID и др.);

- телекоммуникационные ресурсы (ресурсы передачи данных).

В свете вышеизложенного, проблема обеспечения живучести на протяжении ЖЦ связана с необходимостью поддерживать приемлемую (желательно – оптимальную) траекторию деградации функциональных системных возможностей по мере подготовки КИС к выводу из эксплуатации и последующего демонтажа.

Живучесть обычно проявляется в условиях накопления в системе близкого к критическому множества отказов, когда возврат к прежним режимам функционирования в силу внутренних и внешних обстоятельств представляется маловероятным. Например, может надолго выйти из строя важное устройство, либо на неопределенный период отказать каналы связи с некоторыми из корпоративных узлов обработки данных (УОД). Живучесть в этом случае может рассматриваться как функциональная надежность КИС в долгосрочном плане, т.е. как динамическое функциональное соответствие системы изменяющимся условиям эксплуатации в течение длительного периода.

В результате закономерно возникают вопросы о том, какая система может считаться более живучей в функциональном плане? Как оценивать качественные и количественные показатели живучести, какие критерии применимы для такой оценки? Какие решения следует принимать службам эксплуатации для повышения степени живучести столь сложных систем, как современные КИС?

В качестве ответа на первый вопрос можно предположить, что, по-видимому, более живучей следует считать систему, которая в наибольшей степени соответствует декларированным целям своего функционирования в условиях постепенного накопления отказов оборудования, моральной деградации ПО, ухудшения условий эксплуатации и воздействия других внешних неблагоприятных факторов на протяжении всего периода ЖЦ.

Далее, из нескольких альтернативных вариантов более живучей может считаться система, способная выдержать на заданном временном отрезке большее число отказов или большее число тяжелых отказов по сравнению с другими аналогами, а также способная функционировать с качеством не ниже заданного в течение более длительного времени и в наиболее небла-

гоприятных условиях. Возможные ответы на другие поставленные вопросы рассматриваются ниже.

Критерии оценки живучести КИС

Учитывая высокую сложность и комплексный характер проблемы анализа надежности и живучести КИС, логично предположить, что произвести оценку этих показателей только на основе одного какого-либо параметра крайне затруднительно. Поэтому в данном случае необходимо включать в рассмотрение некие интегральные многофакторные (многокритериальные) показатели, включающие как количественные, так и качественные характеристики, которые могут использоваться для формальной оценки ее живучести.

Функциональную живучесть КИС целесообразно определять посредством оценки показателей качества ее функционирования в условиях возникновения отказов в процессе эксплуатации на протяжении всего жизненного цикла. В такой постановке качество выполнения системой своих априорно заданных функций можно оценивать, например, на основе следующих основных характеристик:

- соответствие КИС целям и задачам корпоративного бизнеса;
- показатели производительности системы и отдельных ее узлов;
- функциональная готовность приложений и данных, связанная с показателями отклика (временем реакции системы);
- качество обслуживания пользователей и приложений (QoS);
- рациональное использование ИКТ-ресурсов, оцениваемое с позиций факторов TCO (Total Cost of Ownership) и ROI (Return of Investment).

При проектировании КИС, обладающих требуемой высокой степенью функциональной надежности и живучести, изначально следует определить соответствующие критерии для оценки различных системных качеств. Эти критерии целесообразно объединять по группам в соответствии с определенными принципами, характеризующими различные конструктивно-технологические аспекты, факторы и механизмы обеспечения живучести КИС. Группы могут формироваться, например, по следующим признакам:

- критерии соответствия системы заданным показателям качества функционирования и/или оценки степени ее функциональной (физической и моральной) деградации;
- критерии для оценки эффективности динамической реконфигурации и перераспределения ИКТ-ресурсов, а также динамики восстановления функциональных возможностей системы после сбоев;
- критерии, характеризующие изменение производительности и реактивности системы при выполнении различных типов приложений в условиях деградации системных ресурсов;

- критерии экономической эффективности использования ИКТ-активов.

К первой группе могут быть отнесены критерии, характеризующие:

- долю физически и морально устаревшего оборудования в составе ИКТ-инфраструктуры КИС (серверных платформ, рабочих станций, сетевых устройств и проч.) по отношению к общему объему парка технических средств с использованием комплексных показателей типа SWaP (Space, Watts and Performance);

- долю морально устаревшего программного обеспечения (ПО) в общем объеме прикладного и общесистемного ПО КИС;

- долю гибкого, переносимого и хорошо масштабируемого сервис-ориентированного ПО по отношению к «монолитным» корпоративным приложениям класса ERP;

- вероятность потери функциональности из-за невозможности выполнения некоторых ответственных приложений на оставшихся ИКТ-ресурсах в условиях выхода из строя части ИКТ-инфраструктуры;

- соответствие характеристик и параметров системы в течение ЖЦ требуемым значениям показателей качества обслуживания типа QoS, заданным в соглашениях по SLA (Service Level Agreement);

- темпы развития/деградации ИКТ-инфраструктуры КИС на протяжении определенного периода или всего жизненного цикла системы;

- нечеткие критерии, характеризующие качество функционирования КИС и степень ее соответствия целям и задачам корпоративного бизнеса, построенные, как правило, на основе экспертных оценок с использованием лингвистических переменных;

- критерии оценки эффективности реинжиниринга бизнес-процессов и степени адаптации прикладного ПО системы к новым бизнес-целям и задачам в процессе модернизации/реконструкции КИС.

Вторая группа объединяет критерии, характеризующие способность эксплуатируемой системы противостоять возможным отказам путем реконфигурации и адаптации к новым условиям функционирования:

- эффективность методов и средств виртуализации ИКТ-ресурсов, а также их консолидации для решения критически важных и ответственных задач;

- эффективность перераспределения ИКТ-активов КИС в целях сохранения максимально возможной степени функциональности системы;

- критерии для оценки адаптационных способностей КИС к новым функциональным задачам и возможным архитектурным перестройкам;

- критерии для оценки скорости адаптации системы к условиям изменяющейся нагрузки;

- критерии оценки степени функциональной деградации вследствие роста требований к качеству функционирования, а также из-за изменения (ухудшения, ужесточения) других условий эксплуатации;

- критерии для оценки динамики (темпов) реконфигурации и адаптации системы, а также длительности восстановления функциональных возможностей.

К третьей группе целесообразно относить вероятностные критерии, позволяющие адекватно оценить:

- вероятность и параметры снижения производительности и реактивности системы при изменении конфигурации вследствие отказов, сбоев или вывода части ИКТ-ресурсов из эксплуатации (временного или постоянного);

- требования к активным системным ИКТ-ресурсам, необходимым для обеспечения значений показателей по производительности и реактивности, адекватных входному трафику и характеру запросов на обслуживание;

- характеристики, связанные с вероятностными оценками удовлетворения системой требований по производительности и реактивности в условиях изменения нагрузки путем оптимизации конфигурации ИКТ-ресурсов;

- вероятностные оценки успешного выполнения ответственных и критически важных приложений в условиях прогнозируемых сбоев, отказов и деградации системных ресурсов.

Наконец, последняя группа содержит критерии, объединенные целью эффективного использования имеющихся в КИС ИКТ-ресурсов, а именно:

- критерии для оценки рациональности вложения средств для модернизации/утилизации ИКТ-ресурсов в связи с изменением внешней бизнес-среды (рыночной конъюнктуры) или принятием принципиально иной корпоративной стратегии ведения бизнеса, что в свою очередь связано с неизбежным переопределением бизнес-целей;

- технико-экономические критерии типа ROI (Return of Investment) и TCO (Total Cost of Ownership) с отслеживанием динамики их изменений на протяжении всего ЖЦ КИС;

- показатели для оценки минимально необходимого уровня текущего финансирования для сохранения функциональных возможностей системы на требуемом уровне вплоть до ее вывода из эксплуатации и последующей утилизации.

Суммируя все вышесказанное, следует отметить, что именно живучесть может рассматриваться как наиболее объективный и адекватный показатель, позволяющий наилучшим образом оценить все аспекты структурно-функциональной надежности СТС, находящейся в постоянно изменяющейся внешней среде и подвергающейся перманентным модернизациям с целью улучшения показателей качества ее функционирования. Действительно, при исследовании живучести акцент делается не на единичные сбои и отказы, вызывающие временную неработоспособность системы, а на ее способность выполнять свои функции в течение длительного периода времени, желательно - на протяжении всего ЖЦ КИС.

Конечной целью исследований в области изучения факторов живучести КИС является построение высоконадежных систем, обладающих необходимыми адаптационными свойствами на случай неблагоприятного изменения внешней ситуации, например, резкого ухудшения условий эксплуатации или изменения целей корпоративного бизнеса.

Формализация задачи анализа факторов живучести КИС

В наиболее общей форме постановка проблемы анализа факторов живучести КИС связана с оценкой показателей качества ее функционирования (ПКФ), производимой либо на протяжении всего ЖЦ данной системы, либо на достаточно продолжительном отрезке этого цикла, чаще всего, - в фазе контролируемой деградации объекта исследования.

Пусть качество функционирования КИС определяется множеством показателей (факторов) $\boldsymbol{Q} = \{q_1, \ldots, q_S\}$, которые должны находиться в пределах установленных значений эксплуатационных характеристик. Обычно эти значения задаются в виде технических требований (ТТ), сформулированных в техническом задании (ТЗ) на данную систему.

Введем ограничения вида

$$q_j \geq q_j^{TT} \quad \text{или} \quad q_j \leq q_j^{TT}, \quad j = \overline{1,S}, \tag{1}$$

которые формализуют условия соответствия ПКФ данного объекта (КИС) определенным в ТЗ техническим требованиям. Здесь параметры q_j^{TT}, $j = \overline{1,S}$, представляют собой предельно допустимые граничные значения соответствующих ПКФ системы.

В наиболее простой форме проблема обеспечения живучести может быть сформулирована как задача минимизации стоимости вложений, необходимых для поддержания функциональных ПКФ системы на требуемом уровне в течение заданного периода времени, т.е. в виде:

$$C = \sum_{i=1}^{\Omega} \sum_{j=1}^{S} c_{ij}(\Delta t_i) \to \min_{[0,T]}, \tag{2}$$

$$q_j^*(\Delta t_i) \geq q_j^{TT}, \; j = \overline{1,l}, \quad q_j^*(\Delta t_i) \leq q_j^{TT}, \; j = \overline{l+1,S}, \tag{3}$$

$$T_{пл} = \sum_{i=1}^{\Omega} \Delta t_i \quad ,$$

где $c_{ij}(\Delta t_i)$ – текущие эксплуатационные затраты на календарно-плановом отрезке $\Delta t_i \in [0, T_{пл}]$, $i = \overline{1,\Omega}$, при которых обеспечивается уровень качества функционирования по j-му ПКФ, удовлетворяющий заданным ТТ; $q_j^*(\Delta t_i)$ - «наихудшее» в смысле выполнения условий ТТ значение j-го ПКФ на отрезке Δt_i.

Отметим, что постановка задачи обеспечения живучести в виде (2), (3) справедлива в основном для сепарабельных функций c_{ij}, когда вложения в некоторый тип ресурса, производимые с целью улучшения одного из показателей качества q_j не влияют на другие ПКФ, а также при условии аддитивности интегральной функции стоимости C.

Здесь целесообразно еще раз подчеркнуть, что надежность и отказоустойчивость – это показатели текущие, ситуационные, оперативные. А функциональная живучесть – показатель агрегированный, долгосрочный, характеризующий изменение возможностей исследуемой системы в течение достаточно продолжительного временного отрезка (периода деградации). Поэтому соответствующие вложения, направленные на обеспечение живучести КИС, необходимо осуществлять равномерно в течение всего планового периода эксплуатации, хотя на практике это далеко не всегда соблюдается.

Количественная оценка живучести сложной системы с учетом ее возможностей противостоять функциональной деградации может быть сделана на основе конкретных метрик, характеризующих потерю функциональности на протяжении некоторого временного периода. При этом возможны различные методические подходы к вычислению этих метрик, например, через количественные оценки относительной способности системы к выполнению критически важных и ответственных задач, с учетом утраты ею части начальных возможностей из-за деградации.

В данной работе для целей анализа живучести и оценки степени деградации КИС предлагается использовать интегральный показатель функциональной живучести Φ, который определяется через средневзвешенную сумму оценок ПКФ в следующем виде:

$$\Phi = \frac{1}{S}\sum_{j=1}^{S} z_j(k), \qquad (4)$$

где значения нормированных показателей $z_j(k)$, $j=\overline{1,S}$, вычисляются как

$$z_j(k) = a_j \frac{q_j^*(k) - q_j^{TT}}{q_j^{TT}},\ j=\overline{1,l},\ \text{для ТТ вида}\quad q_j \ge q_j^{TT} \qquad (5а)$$

или

$$z_j(k) = a_j \frac{q_j^{TT} - q_j^*(k)}{q_j^{TT}},\ j=\overline{l+1,S},\ \text{для ТТ вида}\quad q_j \le q_j^{TT}. \qquad (5б)$$

Здесь a_j - весовой коэффициент, характеризующий степень значимости j-го ПКФ для интегральной оценки качества функционирования системы в целом; k - количество накопленных отказов в системе за рассматриваемый период времени (в том числе с учетом восстановлений).

Очевидно, что если для всех заданных ПКФ в рассматриваемый период времени имеет место выполнение требований ТТ вида (1), то

$$\min_j \; z_j(k) \geq 0, \quad j = \overline{1,S} \,,$$

и, следовательно, значение предложенного интегрального показателя Φ будет не ниже некоторой критической нижней границы $\Phi_{кр}$. Ее конкретное значение может задаваться изначально при определении функциональных возможностей системы на определенный период эксплуатации, также как и начальное значение интегрального показателя Φ. Примерный вид графика деградации функциональных возможностей системы представлен на рис. 1а,б.

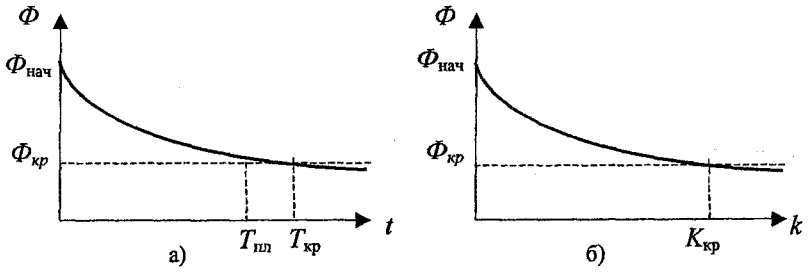

Рис. 1. Деградация функциональных возможностей невосстанавливаемой системы: а) с течением времени; б) с накоплением числа отказов k

В общем виде количественная оценка $S_Д$ степени деградации функциональных возможностей в рамках постановки вида (4) может быть выполнена, например, согласно формуле:

$$S_Д = \frac{\Phi_{нач} - \Phi_{тек}}{\Phi_{нач}} \times 100\% = \frac{\Phi_{утр}}{\Phi_{нач}} \times 100\% \,, \qquad (6)$$

где $\Phi_{нач}$ - количественная оценка начальных (т.е. полных) функциональных возможностей системы с учетом весовых коэффициентов важности ПКФ; $\Phi_{тек}$, $\Phi_{утр}$ – то же для текущих (наличествующих) и утраченных функциональных возможностей соответственно.

В качестве канонической модели для исследования и оценки показателей живучести часто рассматривается следующая [5-7]. Имеется сложная вычислительная система, состоящая из Λ однотипных многофункциональных модулей (МФМ), где изначально каждый модуль способен выполнять априорно заданный набор функций $F = \{f_1, f_2, ..f_L\}$. Однако в процессе эксплуатации в результате частичной потери функциональных возможностей из-за сбоев и отказов некоторые МФМ могут оказаться в состоянии выполнять только ограниченное подмножество из первоначального набора функций F. В результате получим систему, состоящую из частично работоспособных элементов. Если модули в системе соединены скоростными

каналами связи, то при необходимости может осуществляться реконфигурация системы, т.е. выполнение некоторых потерянных функций будет возложено на другие МФМ. При этом не исключены ситуации, когда определенные функциональные возможности будут утеряны на достаточно длительный период или окончательно. В таком случае под функциональным отказом следует понимать неспособность системы выполнять некоторую отдельно взятую функцию из F.

Предполагается также, что каждый модуль может одновременно выполнять только одну функцию из заданного набора, причем в случае необходимости для получения гарантированного результата ту же функцию параллельно могут выполнять еще несколько резервных модулей.

С учетом того, что практически одновременно могут поступать запросы на выполнение всех типов функций из F, количество МФМ в системе должно быть не меньше L, а с учетом резервирования – еще больше, т.е.

$$\Lambda \geq \sum_{i=1}^{L} r_i ,$$

где r_i – коэффициент резервирования МФМ, предназначенного для выполнения i-й функции.

При этом два или несколько однотипных запросов, претендующих на выполнение одной и той же функции, в системе одновременно выполняться не могут, они могут только становиться в очередь на обслуживание.

В тех или иных модификациях эта модель вычислительной архитектуры, рассматриваемая как система массового обслуживания (СМО), неоднократно использовалась для анализа надежности и живучести в условиях различных законов обслуживания, типов потоков заявок, характеристик распределения временных интервалов наработки на отказ/восстановления, в том числе с учетом возможных отказов в каналах связи [7-9].

Живучесть в рамках подобных моделей может исследоваться в самых разных аспектах, например, путем оценки некоторых частных показателей деградации системы, либо через вероятностные оценки успешности перераспределения ресурсов в случае отказа части МФМ. Возможны и другие варианты. Количественное исследование живучести для подобной модели может производиться также с учетом возможных потерь функциональности на основе вероятностных оценок [9-11].

Недостатки подобной модели достаточно очевидны. Во-первых, реальные вычислительные системы обычно состоят из гетерогенных компонентов, назначение, функции и возможности которых сильно различаются. И хотя, например, размещаемые в одной стойке-шасси «лезвийные» серверы близки к рассмотренной модели однотипных МФМ, однако и они могут иметь разные характеристики по объему памяти, производительности и т.д. Во-вторых, в реальных условиях поступающие запросы на обслуживание таковы, что потребности в одних и тех же ИКТ-ресурсах возникают

постоянно, поэтому архитектура и возможности современных КИС должны позволять обрабатывать одновременно сотни однотипных запросов, даже при необходимости многократного резервирования при выполнении критически важных задач. В третьих, как показывает рассмотренный выше перечень критериев оценки живучести, современные КИС являются гораздо более сложным и многогранным явлением, чем можно представить в рамках рассмотренной канонической схемы. Следовательно, для полноценного анализа их живучести необходимы более содержательные модели.

Модели для исследования живучести КИС (подсистемы front-end)

Рассмотрим конкретные примеры формализованных моделей, предназначенных для анализа надежности и живучести КИС. Для этих целей целесообразно применять подходы на основе теории вероятностей, теории массового облуживания и исследования операций, а также методы исследования моделей с использованием аппарата конечных цепей Маркова, включая цепи с доходами, и стохастических сетей Петри [10-16]. Примем также в качестве исходных базовых положений следующие:

- построение и анализ моделей для оценки живучести КИС целесообразно осуществлять отдельно для подсистем класса front-end и подсистем back-end в силу различия решаемых ими функциональных задач, особенностей архитектурных решений и нюансов технической реализации;
- исследование живучести должно производиться неразрывно с анализом вероятной нагрузки подсистем КИС в рамках одних и тех же моделей;
- оценку показателей живучести следует проводить относительно некоторого заданного календарно-планового периода эксплуатации системы.

Пусть в составе КИС функционирует корпоративный Web-узел, выполняющий функции подсистемы front-end и построенный в виде множества web-серверов, которые реализованы на базе стойки серверов-«лезвий» (blade server). Будем считать, что Web-узел состоит из N «лезвий», функционирующих как автономные серверы со своими процессорными модулями и полями памяти, и что в системе в целом имеет место пуассоновский поток отказов с интенсивностью λ. При этом часть серверов может выполнять полезную нагрузку, а остальные - находиться в состоянии «горячего» и/или «теплого» резерва, т.е. в данном случае используется так называемая «схема с нагруженным резервом» [10, 12, 21].

Если система в силу определенных обстоятельств, например, из-за существующих финансовых ограничений, не подвергается профилактике и хотя бы частичному ремонту, то она неуклонно деградирует и, так или иначе, должна выводиться из эксплуатации. Если же система должна поддерживаться в рабочем состоянии с заданным уровнем качества функционирования, то производится постоянное восстановление отказавших серверов. В особо ответственных случаях может осуществляться оперативное

восстановление функций отказавших серверных «лезвий» путем временного изъятия их из стойки и замены на резервные с последующим ремонтом изъятых серверов или без него. Этот случай известен в практике как схема «с ненагруженным резервом» (с восстановлением и/или без него).

В случае невосстанавливаемой деградирующей КИС функциональную живучесть ее подсистемы front-end на заданный планово-календарный период $T_{пл}$ можно оценивать через вероятность того, что на отрезке $[0,T_{пл}]$ число отказавших серверных «лезвий» (имеются в виду фатальные отказы, при которых отказавшие устройства не заменяются из-за отсутствия работоспособных резервных узлов), будет не больше $K_{кр}$. Как известно, вероятность наступления не более K событий за период $(0,t)$ для потока с интенсивностью λ может быть определена по формуле [20]:

$$P_K(t) = \sum_{k=0}^{K} P_k(t) = e^{-\lambda t} \sum_{k=0}^{K} \frac{(\lambda t)^k}{k!} .$$

Отсюда может быть найдена вероятностная оценка жизнеспособности подсистемы на конец рассматриваемого календарного периода $T_{пл}$ для состоящей из N «лезвий» серверной стойки-шасси, рассматриваемой как горизонтально масштабируемый ИКТ-ресурс. Очевидно, что в данном случае для анализа живучести необходимо прежде всего выявить границу $K_{кр}$, после достижения которой в результате накопления отказов не соблюдается выполнение требований ТЗ. Определение $K_{кр}$ связано с практическими замерами ПКФ подсистемы front-end на реальном объекте с последующей оценкой показателей производительности [10, 11].

Величина $K_{кр}$ может быть установлена эмпирическим путем на основе натурных экспериментов, в ходе которых из стойки-шасси производится последовательное изъятие некоторого количества серверных «лезвий» как имитация их отказов с последующими замерами реальных ПКФ и/или их вычислением по результатам экспериментов. Граничное значение $K_{кр}$ определяется ситуацией, когда после очередного изъятия «лезвий» реальные ПКФ подсистемы перестают отвечать условиям ТТ вида (1), при этом в стойке остается функционировать $N_{кр}$ устройств. Таким образом, живучесть рассматриваемой невосстанавливаемой подсистемы обеспечивается при условии:

$$N_{кр} \geq N - K_{кр}.$$

Значительно больший практический интерес представляют модели для оценки живучести систем с восстановлением вычислительных мощностей, в которых возможен возврат системных кондиций до значений ПКФ, отвечающих ТТ. Для формализации задач построения и анализа подобных моделей будем также использовать аппарат теории массового обслуживания и конечных цепей Маркова [10, 20-22].

В этих условиях подсистема front-end, построенная на базе стойки-шасси web-серверов из N «лезвий», может быть представлена в виде СМО,

содержащей N параллельно включенных обслуживающих устройств (ОУ), обрабатывающих поступающие от пользователей заявки на обслуживание (рис. 2а). Будем также считать, что подсистема обладает некоторым входным буфером памяти объемом m для хранения очереди ожидающих обслуживания заявок, что позволяет избежать их потерь в случае, если все остающиеся в рабочем состоянии серверы в стойке будут заняты обработкой ранее поступивших заявок.

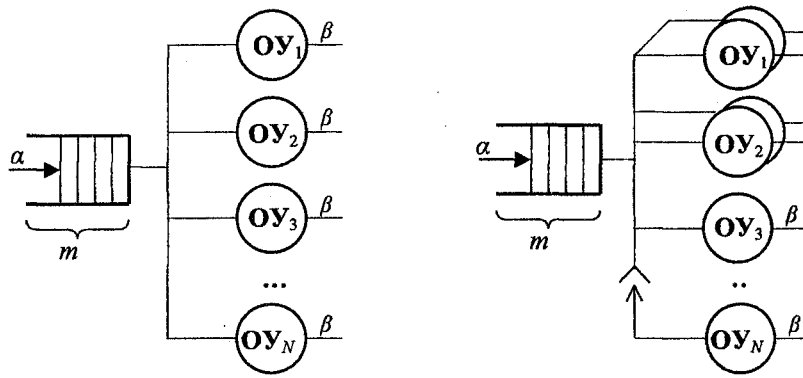

Рис. 2. Модель СМО для подсистемы front-end с входным буфером заявок: а) без резервирования; б) с «горячим» и «холодным» резервированием

Введем в рассмотрение следующие состояния подсистемы front-end: E_0 – в системе нормально функционируют все N серверных «лезвий»; E_1 – один из серверов находится в неработоспособном состоянии, осуществляется его ремонт; E_2 – два сервера в состоянии отказа; и далее E_N – все имеющиеся в корпоративной стойке серверные «лезвия» находятся в ремонте, производится их восстановление. Если считать, что каждый из серверов в системе испытывает пуассоновский поток отказов с интенсивностью λ, а восстановление каждого из них осуществляется с интенсивностью μ, то граф переходов состояний системы в виде конечной марковской цепи может быть представлен, как показано на рис. 3. Вероятности нахождения подсистемы в каждом из состояний $E_0, E_1 \ldots E_N$ в установившемся режиме обозначим как $p_0, p_1, \ldots, p_N$ соответственно.

Как известно, подобная марковская модель переходов состояний сложной системы называется схемой «гибели-размножения» [21]. Введем

также в рассмотрение приведенную интенсивность потока отказов, которая определяется как $\rho = \lambda/\mu$. Ясно при этом, что по мере увеличения числа отказавших серверов общая производительность подсистемы падает, а по мере их восстановления – растет.

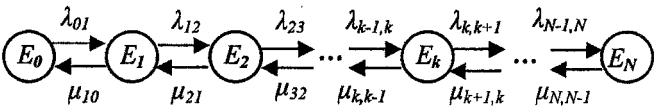

Рис. 3. Марковская модель переходов состояний для восстанавливаемой подсистемы «front-end»

Тогда базовая модель для расчета множества вероятностных характеристик, определяющих степень живучести подсистемы front-end, рассчитывается через так называемые финальные вероятности состояний системы на основании формул Эрланга [20-22].

С учетом того, что $\lambda_{01} = N\lambda$; $\lambda_{12} = (N-1)\lambda$; $\lambda_{23} = (N-2)\lambda$; ... $\lambda_{N-1,N} = \lambda$, а также, что $\mu_{10} = \mu$; $\mu_{21} = 2\mu$; $\mu_{32} = 3\mu$; ...; $\mu_{N,N-1} = N\mu$, она (модель) может быть представлена в следующем виде:

$$
\begin{aligned}
& p_0 = (1 + N\rho + \frac{N(N-1)}{2!}\rho^2 + \frac{N(N-1)(N-2)}{3!}\rho^3 + ... + \rho^N)^{-1}; \\
& p_1 = N\rho p_0; \quad p_2 = \frac{N(N-1)\rho^2}{2!}p_0; \quad p_3 = \frac{N(N-1)(N-2)\rho^3}{3!}p_0; \\
& \ldots \\
& p_k = \frac{N!\rho^k}{(N-k)!\,k!}p_0; \quad \ldots p_N = \rho^N p_0; \\
& \sum_{i=0}^{N} p_i = 1.
\end{aligned}
\qquad (7)
$$

Отсюда с использованием системы уравнений (7), которая описывает финальные вероятности всех возможных состояний системы в рамках данной модели, искомая вероятность того, что подсистема переднего плана front-end будет в той или иной мере оставаться жизнеспособной, т.е. в установившемся режиме не менее (N-K) серверных «лезвий» в стойке будут работоспособны (при $K \le K_{кр}$) вычисляется как:

$$P_{ЖС} = \sum_{i=0}^{K} p_i = (1 + N\rho + \frac{N(N-1)\rho^2}{2!} + ... + \frac{N!\rho^K}{(N-K)!K!})p_0 . \qquad (8)$$

По мере снижения вследствие отказов числа работоспособных «лезвий» в стойке производительность и другие технические характеристики web-узла неуклонно деградируют, однако подсистема в целом считается работоспособной, пока поддерживаются на должном уровне ее показатели качества функционирования. При этом время от времени некоторые ПКФ могут отклоняться за пределы ТТ, но в процессе восстановления подсистема возвращается в предписанные рамки, обозначенные в ТЗ.

Поскольку в восстанавливаемых подсистемах КИС нарушение условий вида (1) может являться временным явлением, важно оценить жизнеспособность таких объектов в среднем за период эксплуатации, т.е. их долгосрочную устойчивость по отношению к влиянию дестабилизирующих факторов. Используя (7) и (8), можно определить среднее значение числа работоспособных серверов $N_{ср}$ в рамках рассмотренной марковской модели. Оно определяется как математическое ожидание дискретной случайной величины (т.е. количества действующих серверов) из выражения:

$$N_{ср} = \sum_{i=0}^{N-1} p_i(N-i) = p_0 \sum_{i=0}^{N-1} \frac{N!}{(N-i)!} \frac{\rho^i}{i!}(N-i). \qquad (9)$$

В рассматриваемой ситуации для обеспечения живучести системы также требуется выполнение условия

$$N_{ср} \geq N - K_{кр} .$$

Для подсистем класса front-end опасность потери функциональности из-за отказов серверных «лезвий» связана, в основном, с невозможностью адекватного обслуживания поступающего по корпоративной сети потока заявок на обслуживание. Это не только приводит к образованию очередей и возможному переполнению входного буфера, но чревато также потерей заявок, либо превышением заданных нормативов их обработки, что в конечном итоге обусловливает определенное ухудшение характеристик качества обслуживания, пусть даже и временное.

В то же время, в реальных КИС живучесть подобных подсистем во многом зависит не только и не столько от числа работоспособных web-серверов, сколько от того, насколько успешно они справляются с фактическим потоком пользовательских заявок на обслуживание. В частности, в системе может быть достаточно много отказавших и находящихся в состоянии восстановления серверов, но при этом и реальная нагрузка может быть не слишком велика, вследствие чего система в целом остается «на плаву», поскольку при невысокой нагрузке ее основные ПКФ остаются в пределах установленных в ТТ нормативов.

Для такого случая оценить параметры снижения качества обслуживания можно на основе модели СМО (рис. 2а), формально описываемой как M/M/N/m [21]. Пусть поток заявок носит пуассоновский характер с интенсивностью α. Каждое ОУ характеризуется интенсивностью обработки заявок (производительностью) β_i, $i=\overline{1,N}$. При необходимости следует также учитывать, что интенсивность входного потока заявок α может меняться во времени, начиная с небольших значений в утренние часы и достигая пиковых значений через 1-2 часа после начала рабочего дня, а затем постепенно снижается. Следовательно, в общем случае поток заявок на достаточно длительном временном интервале не является стационарным.

Для простоты рассуждений будем далее считать, что в модели СМО все обслуживающие устройства (ОУ) характеризуются одинаковой производительностью, т.е.

$$\beta_i = \beta, \ i=\overline{1,N}.$$

Приведенная интенсивность обработки заявок определяется как

$$\psi = \alpha/\beta.$$

С учетом имеющегося в подсистеме входного буфера заявок объемом m марковская модель переходов состояний в зависимости от нахождения в ней определенного количества заявок на обслуживание может быть представлена, как показано на рис. 4. Согласно модели рассматриваемая СМО может находиться в одном из следующих состояний: S_0 – в подсистеме отсутствуют заявки, все ОУ свободны; S_1 – в СМО присутствует одна заявка, одно из ОУ занято ее обработкой; S_2 – два ОУ заняты обработкой, остальные свободны; S_N – в системе присутствует N заявок, все имеющиеся ОУ заняты, причем входной буфер заявок пуст; и далее S_{N+m} – все имеющиеся ОУ заняты, системный буфер полностью заполнен заявками, ожидающими в очереди на обслуживание. Финальные вероятности нахождения системы в этих состояниях (при условии стационарности входного потока заявок) обозначим как $p_0, p_1, \ldots, p_N, \ldots, p_{N+m}$ соответственно.

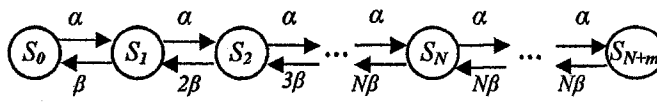

Рис. 4. Модель переходов состояний для СМО с входным потоком заявок α и входным буфером объемом m (рис. 2а)

Тогда на основании формул Эрланга для модели переходов состояний (рис. 4) эти вероятности могут быть определены следующим образом:

$$p_0 = (\sum_{i=0}^{N} \frac{\psi^i}{i!} + \frac{\psi^N}{N!} \sum_{i=1}^{m} (\frac{\psi}{N})^i)^{-1};$$

$$p_1 = \psi p_0; \quad p_2 = \frac{\psi^2}{2!} p_0; \quad p_3 = \frac{\psi^3}{3!} p_0; \cdots; \quad p_N = \frac{\psi^N}{N!} p_0; \qquad (10)$$

$$p_{N+1} = \frac{\psi^N}{N!} \frac{\psi}{N} p_0; \quad \cdots; \quad p_{N+m} = \frac{\psi^N}{N!} (\frac{\psi}{N})^m p_0;$$

$$\sum_{i=0}^{N} p_i = 1.$$

Наиболее важными характеристиками для модели переходов состояний СМО, описываемой системой уравнений (10), являются следующие:

- среднее число загруженных обработкой заявок ОУ (серверных «лезвий»)

$$n_{cp} = \sum_{i=1}^{N} i p_i + N \sum_{i=N+1}^{N+m} p_i = p_0 (\sum_{i=1}^{N} i \frac{\psi^i}{i!} + \frac{\psi^N}{(N-1)!} \sum_{i=1}^{m} (\frac{\psi}{N})^i);$$

- средняя длина очереди заявок, ожидающих обслуживания в буфере

$$D_{cp} = \sum_{i=1}^{m} i p_{N+i} = p_0 \frac{\psi^N}{N!} \sum_{i=1}^{m} i (\frac{\psi}{N})^i;$$

- среднее число заявок, находящихся в рассматриваемой подсистеме

$$\Theta = n_{cp} + D_{cp}.$$

Поскольку в данной СМО одновременно имеет место и поток отказов, можно констатировать, что возможны ситуации, когда некоторая часть ОУ, например, K, будет неработоспособна. А значит, меняется не только формализм описания системы, т.е. она должна быть представлена, как M/M/N-K/m, но и ухудшаются соответствующие характеристики производительности и некоторые другие ПКФ.

Так как в данном случае имеют место независимые пересекающиеся потоки событий (отказы ОУ и поступление заявок на обслуживание), то для анализа протекающих в системе процессов целесообразно построить марковскую модель в двух измерениях, как показано на рис. 5. В ней состояния E_0, E_1,..., E_N , представленные в самой верхней строке, соответствуют, как и ранее, различному числу отказавших ОУ, но при отсутствии в системе каких-либо заявок на обслуживание.

Каждый столбец модели соответствует некоторому фиксированному количеству работоспособных ОУ в системе, а каждая строка – некоторому количеству поступивших заявок, обрабатываемых или ожидающих в очереди. При этом, если рассматривать модель сверху вниз, то с каждым переходом на последующую строку модели в ней увеличивается число нахо-

дящихся в СМО заявок, а переходы слева направо отражают рост числа отказавших устройств.

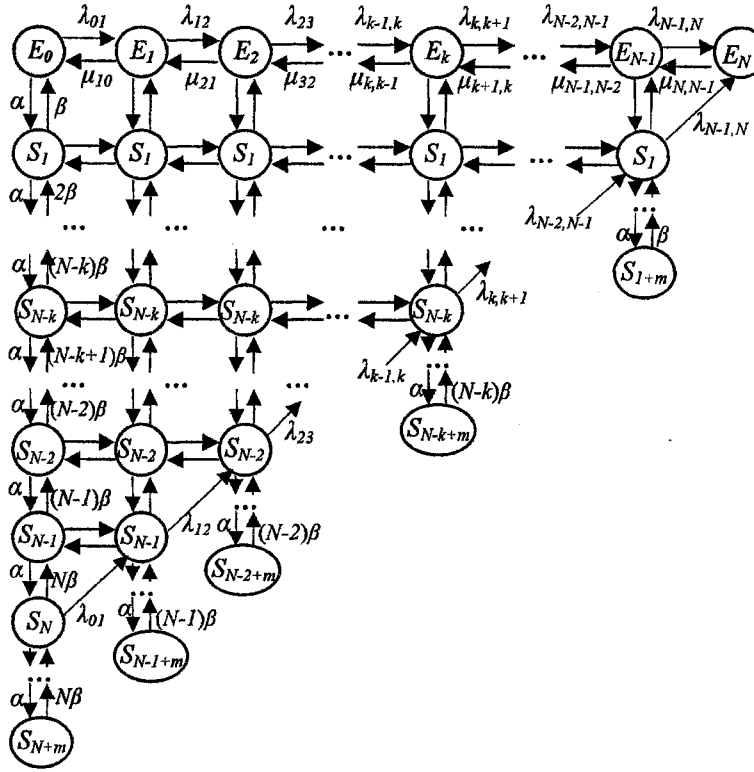

Рис. 5. Модель переходов состояний для СМО с отказами/восстановлением ОУ

Следует также отметить, что «косые» стрелки под углом 45° на рисунке соответствуют переходам состояний, связанным с отказами ОУ, что одновременно приводит к потере обрабатываемых ими в этот момент заявок. Для простоты такие переходы показаны только для состояний с полной загрузкой всех рабочих ОУ. Обратные «косые» переходы, связанные с одновременным восстановлением работоспособности ОУ и возобновлением прерванного процесса обработки невозможны, точнее, возможны только при наличии кэш-памяти обрабатываемых заявок или в схемах с резервированием.

Опираясь на данную двухмерную модель переходов состояний, можно сформулировать в качестве условий живучести для системы с восстановлением и стационарным потоком заявок следующие положения:

- среднее число занятых обработкой заявок ОУ не должно превышать среднего числа работоспособных серверов согласно (9), т.е.

$$N_{cp} \geq n_{cp}; \tag{11}$$

- среднее число заявок, находящихся в рассматриваемой СМО, не должно быть больше числа работоспособных ОУ плюс объем буфера, т.е.

$$N_{cp} + m \geq \Theta; \tag{12}$$

- среднее время пребывания заявки в СМО с отказами/восстановлением ОУ не должно превышать установленных в ТТ значений и т.д.

Вместе с тем, следует учитывать, что приведенные выше рассуждения и положения, строго говоря, не вполне корректны с математической точки зрения, поскольку в рассмотренной системе нельзя однозначно говорить о независимости «вертикальных» и «горизонтальных» потоков событий. Действительно, финальные вероятности, рассчитанные согласно системе уравнений (10), а также соотношения (11) и (12), вообще говоря, зависят от числа работоспособных серверов в системе. Они могут вычисляться таким образом, если это число остается неизменным вплоть до достижения стационарности потока заявок, т.е. если $\lambda << \alpha$ и $\mu << \beta$.

Данная модель может использоваться не только в целях аналитических исследований различных факторов живучести, но и для имитационного моделирования происходящих в СМО процессов и выявления на основе полученных результатов интересующих пользователей аспектов. Потоки событий при моделировании могут иметь любые значения α, β, λ, μ [14-16]. Путем имитационного моделирования могут исследоваться также гораздо более сложные случаи, когда в системе будут представлены варианты с резервированием ОУ, «горячим» или «холодным», т.е. с нагруженным и ненагруженным резервом. Для таких схем соответствующие марковские модели могут быть трехмерными или даже r-мерными. Представляется полезным исследовать путем моделирования также такой вполне реальный случай, когда для различных ПКФ значения $K_{кр}$ будут существенно отличаться друг от друга. В такой ситуации необходимо создавать специальные многомерные модели для анализа на наихудший случай.

Представляют интерес и иные методы построения моделей, а также технологические приемы для исследования показателей живучести СТС. Так, например, в работах [13-18] для анализа характеристик функциональной живучести (performability) постепенно деградирующих сложных систем применяются усредненные оценки вида:

$$\widetilde{\phi}_j = \sum_{i=0}^{N} \phi_{ij} p_i \ ,$$

где ϕ_{ij} – значение j-го показателя (функции) качества, вычисленное для i-го состояния системы из $E_0, E_1, \ldots, E_N$; p_i - вероятность нахождения системы в состоянии E_i в установившемся режиме.

Тогда, принимая $q_j^*(k) = \widetilde{\phi}_j$, можно воспользоваться предложенными выше соотношениями (4), (5а), (5б) и (6) для средневзвешенной оценки Φ и $S_Д$ как интегральных показателей качества функционирования СТС.

Стоит отметить, что в подсистемах front-end могут возникать и другие процессы, к которым применимы марковские модели, например, влияние дестабилизирующих внешних воздействий. К таковым относятся, например, межсетевые атаки, имеющие целью создать избыточную нагрузку на web-узел для провоцирования неисправностей типа DoS (Deny of Service), т.е. «отказ в обслуживании» [17, 19].

Не менее важный аспект проблемы живучести связан с тем, что с течением времени любая автоматизированная система, даже без накопления отказов, перестает в полной мере отвечать возрастающим потребностям пользователей и подлежит либо серьезной реконструкции, либо постепенному выводу из эксплуатации при минимальных затратах на поддержание необходимого уровня функциональности в течение этого периода. Для КИС такая деградация прежде всего связана с «моральным» старением прикладного ПО (приложений) в подсистемах класса back-end.

Подходы к анализу живучести подсистем back-end

Для подсистем класса back-end оценку живучести наиболее целесообразно осуществлять с точки зрения их способности выполнять критически важные приложения. Желательно рассмотреть три ключевых аспекта, определяющих живучесть КИС с точки зрения реализуемости приложений: функциональность, готовность, масштабируемость.

В свете вопроса о функциональной живучести прикладного ПО целесообразно, в зависимости от степени его важности и ответственности, выделить следующие категории (виды) приложений:

- критически важные;
- ответственные (приоритетные);
- рутинные (неприоритетные);
- офисные;
- фоновые.

Очевидно, что в системах с высокой степенью живучести при выходе из строя в результате отказов части оборудования должна сохраняться безусловная гарантированная реализуемость приложений первой группы и высокая вероятность успешного выполнения как можно большего числа

приложений из второй группы в соответствии с заданными значениями показателей качества. Иначе говоря, постепенная деградация технических характеристик аппаратных средств живучей системы не должна приводить к функциональным последствиям неотвратимого характера с точки зрения выполнения наиболее ответственного прикладного ПО. Система, где не могут выполняться критически важные задачи, не соответствует глобальным корпоративным целям и не может считаться жизнеспособной.

Предложенный выше подход, в основе которого лежат интегральные оценки (4)-(6) в данном случае следует модифицировать таким образом, чтобы оценки функциональности рассчитывались через количество критически важных и ответственных приложений, остающихся в работоспособном состоянии (с учетом весовых коэффициентов, учитывающих их значимость).

Можно также количественно оценивать функциональную живучесть подсистем КИС класса back-end по соотношению числа критически важных и ответственных приложений, адекватно функционирующих в заданном режиме (т.е. в соответствии с заданными значениями QoS), к общему их числу, планируемому к выполнению в некоторый определенный период времени в условиях постоянной деградации характеристик производительности всей системы в целом.

Далее, следующий фактор, характеризующий живучесть подсистем класса back-end, т.е. готовность, обычно рассматривается в виде двух составляющих: операционной и информационной. Операционная готовность опирается на технологии объемно-календарного планирования и характеризует вероятность наличия в подсистеме достаточного количества свободных ИКТ-ресурсов, необходимых для успешного выполнения критически важных приложений на момент их поступления в заданном календарно-плановом периоде. Это достаточно сложная и неоднозначная проблема, которая требует отдельного рассмотрения вне рамок данной статьи.

Информационная готовность подразумевает наличие в подсистеме необходимых информационных ресурсов, хранящихся в базах и банках данных (БД и БнД), которые в необходимый момент могут быть найдены, извлечены и использованы соответствующими приложениями, в том числе критически важными. Показатели информационной готовности обычно характеризуют оперативность и быстродействие системных механизмов поиска и извлечения данных. Другие характеристики такого вида готовности связаны с надежностью хранения данных и непротиворечивостью при восстановлении. Некоторые аспекты этой проблемы, связанные с репликацией ресурсов БД и БнД рассмотрены нами ранее в [12].

Наконец, такой фактор как масштабируемость связан с оценкой реализуемости ряда сложных приложений в неоднородных (гетерогенных) территориально-распределенных информационно-вычислительных средах (ИВС), характерных для КИС. Свойства интероперабельности, мобильно-

сти и переносимости, характерные для хорошо масштабируемых приложений, наиболее оптимальным образом могут быть реализованы в рамках сервис-ориентированных архитектур.

В заключение данного раздела следует отметить, что опасность функциональной деградации КИС в «моральном» отношении из-за ее несоответствия стратегии и целям корпоративного бизнеса вследствие изменения экономических условий существования (рыночной конъюнктуры) чревата невыполнением условий ТТ даже при полной исправности технических средств системы. Такая ситуация может успешно преодолеваться только путем разработки и/или приобретения нового перспективного прикладного ПО с последующей инсталляцией и адаптацией к существующей в рамках КИС общесистемной ИВС.

Механизмы и технологии обеспечения живучести КИС

Основываясь на имеющемся опыте проектирования, создания и внедрения ИС корпоративного масштаба, можно сформулировать следующие принципы организации (пути реализации) проектных аспектов с целью построения КИС, обладающих необходимой степенью живучести:

- применение новых перспективных архитектурных решений КИС, включая территориально-распределенные центры обработки данных, кластеризацию и сервис-ориентированные архитектуры;
- разумное использование как структурной, так и функциональной избыточности устройств обработки и передачи данных, а также телекоммуникационных сетей;
- многофункциональный характер используемых аппаратно-программных модулей, включая «серверные лезвия», их конструктивная и технологическая взаимозаменяемость в режиме с «горячим» и «холодным» резервированием;
- возможность оперативного перераспределения решаемых задач с балансировкой нагрузки между конгломератами ИКТ-ресурсов и их масштабированием;
- эффективная организация вычислительных процессов, применение децентрализованных принципов управления узлами обработки данных и транспортной средой;
- применение специальных механизмов и процедур (например, создание динамических системных доменов), обеспечивающих выполнение критически важных приложений с гарантированным качеством и в строго заданные сроки;
- применение адаптационных механизмов обеспечения живучести в процессе ЖЦ.

В качестве основы для построения перспективных КИС следует ориентироваться на сервис-ориентированные архитектуры SOA (Service-

Oriented Architecture). Основное преимущество данной архитектуры состоит в возможностях построения гибких адаптируемых систем, способных к оперативным функциональным перестройкам и постоянному развитию на протяжении всего жизненного цикла КИС.

Сервисы в рамках SOA являются средствами реализации распределенных масштабируемых прикладных программных компонентов КИС, которые обеспечивают выполнение предписанных им функций обработки данных и обмениваются информацией посредством асинхронного обмена сообщениями. Архитектурная модель SOA обеспечивает быструю сборку распределенных программных объектов в единую среду исполнения за счет использования соответствующих инструментальных средств.

Адаптационные технологии реализации живучих КИС

Конструктивные и технологические особенности построения живучих КИС, обеспечивающие сохранение их функциональных возможностей на протяжении всего жизненного цикла системы, являются предметом исследований для последующих генераций новых поколений подобных систем. Одним из наиболее важных аспектов живучести КИС являются адаптационные возможности систем подобного класса, проявляющиеся в условиях изменения среды эксплуатации.

Адаптация на системном уровне определяется, прежде всего, способностью системы вырабатывать правильную стратегию поведения в связи с изменением условий существования (внешних и внутренних факторов). Адаптация на уровне системных средств и middleware может быть представлена, например, специальными аппаратно-программными механизмами типа кластеризации и динамического перераспределения (балансировки) нагрузки [24-26].

О методике оценки живучести КИС

Методика оценки живучести КИС на протяжении ее ЖЦ может быть разработана с использованием рассмотренных выше критериев и построена на основе подхода, изложенного в [25]. При этом в качестве исходных моментов для оценки степени развития/деградации функциональных возможностей системы используются многокритериальные векторные постановки с применением различных частных критериев, включая нечеткие. Разница с методикой, предложенной в [25], будет заключаться только в противоположном направлении соответствующего вектора развития сложной системы (в данном случае – ее деградации). Подобная постановка проблемы позволяет выбрать рациональные варианты модернизации системы в условиях ухудшения условий эксплуатации, например, из-за роста на-

грузки на серверы вследствие резкого повышения интенсивности потоков заявок на обработку пользовательских запросов.

Для целей анализа живучести необходимо производить соответствующие попарные сравнения альтернатив на базе экспертных оценок, в том числе нечетких с использованием лингвистических переменных. Под альтернативами понимаются возможные варианты архитектурных изменений КИС, связанных, например, с постепенной деградацией узлов обработки и хранения данных, а также телекоммуникационной инфраструктуры вследствие физического и морального износа, приводящей к снижению функциональных возможностей системы. Следовательно, должны быть определены соответствующие метрики, которые могут рассчитываться с целью исследования численных характеристик ПКФ КИС и степени их деградации. При этом возможна привязка количественных и качественных ПКФ к конкретным фрагментам корпоративной ИКТ-инфраструктуры.

Заключение

Создание крупномасштабных КИС, обладающих необходимыми показателями надежности и функциональной живучести, остается одной из актуальных задач системотехники в области ИКТ. Для ее решения необходимо изучение как теоретических аспектов управления показателями качества и надежности разрабатываемых СТС, так и накопление достаточного фактического материала в виде эмпирических данных в данной области. В более общем плане речь, по-видимому, следует вести о вопросах теории и практики управляемой деградации сложных технических систем, поддерживающих адекватный уровень своей функциональной готовности вплоть до полной утилизации в конце жизненного цикла.

Литература

1. Надежность технических систем /Справочник под ред. И. А. Ушакова. – М.: Радио и связь, 1985.

2. Надежность автоматизированных систем управления /Справочник под ред. Я.А. Хетагурова. – М.: Высшая школа, 1988.

3. Дружинин Г.В. Надежность автоматизированных систем. – М.: Энергоатомиздат, 1986.

4. Гуляев В.А., Додонов А.Г., Пелехов С.П. Организация живучих вычислительных структур. – Киев: Наукова думка, 1982.

5. Черкесов Г.Н. Методы и модели оценки живучести сложных систем. – М.: Знание, 1987.

6. Додонов А.Г., Кузнецова М.Г., Горбачик Е.С. Введение в теорию живучести вычислительных систем. – Киев: Наукова думка, 1990.

7. Богатырев В.А. Отказоустойчивость и сохранение эффективности функционирования многомагистральных распределенных вычислительных систем // Информационные технологии. - 1999. - №9. - С.44-48.

8. Богатырев В.А. К оценке отказоустойчивости вычислительных систем с учетом размещения функциональных ресурсов // Информационные технологии. - 2000. - №7. - С.9-13.

9. Богатырев В.А. Надежность отказоустойчивых вычислительных систем реального времени, компонуемых из многофункциональных модулей // Информационные технологии. - 2000. - №10. - С.11-16.

10. Феррари Д. Оценка производительности вычислительных систем. – М.: Мир, 1981.

11. Хорошевский В.Г. Инженерный анализ функционирования вычислительных машин и систем. – М.: Радио и связь, 1987.

12. Зиновьев П.А., Моисеев В.С., Мейко А.В. Модели для оценки надежности архитектурных решений корпоративных систем хранения данных // Исследования по информатике. Вып. 9. - Казань: Отечество, 2005. - С.93-102.

13. Meyer J.F. On Evaluating the Performability of Degradable Computing System // IEEE Trans. on Computers. - 1980. - Vol. C-29. - № 8. – P.720-731.

14. Juneja S., Shahabuddin P. Fast Simulation of Markovian Reliability/Availability Models with General Repair Policies // Proc. 22-nd IEEE Int'l Symposium on Fault-Tolerant Computing. - Boston, MA, 1992. – P.150-159.

15. Catania V., Puliafito A., Vita L. A Model for Performance Evaluation for Gracefully Degrading Systems // Computer Journal. - 1993. - Vol. 36. - № 2. – P.177-185.

16. Jones A. The Challenge of Building Survivable Information-Intensive Systems // IEEE Computer. - 2000. - Vol. 33. - № 8. – P.39-43.

17. Haverkort B.R., Marie R., Rubino G., Trivedi K.S. (Eds). Performability Modelling: Techniques and Tools. – Chichester, England: John Wiley & Sons, 2001.

18. Nicola V.F., Shahabuddin P., Nakayama M. Techniques for Fast Simulation of Models of Highly Dependable Systems // IEEE Trans. on Reliability. - 2001. - Vol. 50. - № 3. – P.246-264.

19. Nicol D.M., Sanders W.H., Trivedi K.S. Model-Based Evaluation: from Dependability to Security // IEEE Trans. on Dependable and Secure Computing. - 2004. - Vol. 1. - № 1. – P.48-65.

20. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1969.

21. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. – М.: Наука, 1987.

22. Клейнрок Л. Вычислительные системы с очередями. – М.: Мир, 1979.

23. Зиновьев П.А. О методологических особенностях системного проектирования корпоративных информационных систем // Исследования по информатике. Вып. 3. - Казань: Отечество, 2001. - С.3-30.

24. Устюгова В.Н. Сравнительный анализ механизмов выравнивания нагрузки с использованием технологии CORBA // Исследования по информатике. Вып. 9. - Казань: Отечество, 2005. - С.155-176.

25. Исмагилов И.И., Зиновьев П.А. Многокритериальная оценка уровня развития сложных технических систем // Исследования по информатике. Вып. 8. - Казань: Отечество, 2004. - С.25-32.

26. Залещанский Б.Д., Чернихов Д.Я. Кластерная технология и живучесть глобальных автоматизированных систем. – М.: Финансы и статистика, 2005.

27. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. - М.: Наука, 2006.