7 КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 681.3

КРИТЕРИИ ОПТИМАЛЬНОСТИ МНОГОУРОВНЕВЫХ ОТКАЗОУСТОЙЧИВЫХ КОМПЬЮТЕРНЫХ СИСТЕМ

В.А. Богатырев, С.В. Богатырев

Рассмотрены варианты формирования частных и обобщенных критериев при решении задачи векторной оптимизации многоуровневых компьютерных систем с учетом надежности, времени обработки запросов и стоимости реализации системы. Показана возможность применения коэффициента сохранения эффективности в качестве объективного обобщенного критерия векторной оптимальности.

Ключевые слова: надежность, отказоустойчивость, векторная оптимизация, критерий оптимальности, коэффициент сохранения эффективности.

Введение

Проектирование отказоустойчивых компьютерных систем связано с решением задачи оптимального резервирования. При многоуровневой структуре компьютерных систем задача оптимизации сводится к определению кратности резервирования узлов каждого уровня, при котором эффективность системы максимальна, а затраты на ее реализацию не превышают допустимого предела. Результат оптимизации во многом зависит от выбора критерия эффективности, формируемого с учетом требований по надежности, производительности (времени ожидания) и стоимости системы. В связи с этим рассматриваются варианты формирования частных и обобщенных критериев эффективности многоуровневых компьютерных систем. Актуальность выбора объекта оптимизации обусловливается тем, что в настоящее время многоуровневая организация широко используется в компьютерных системах, в том числе кластерной архитектуры (рис. 1).

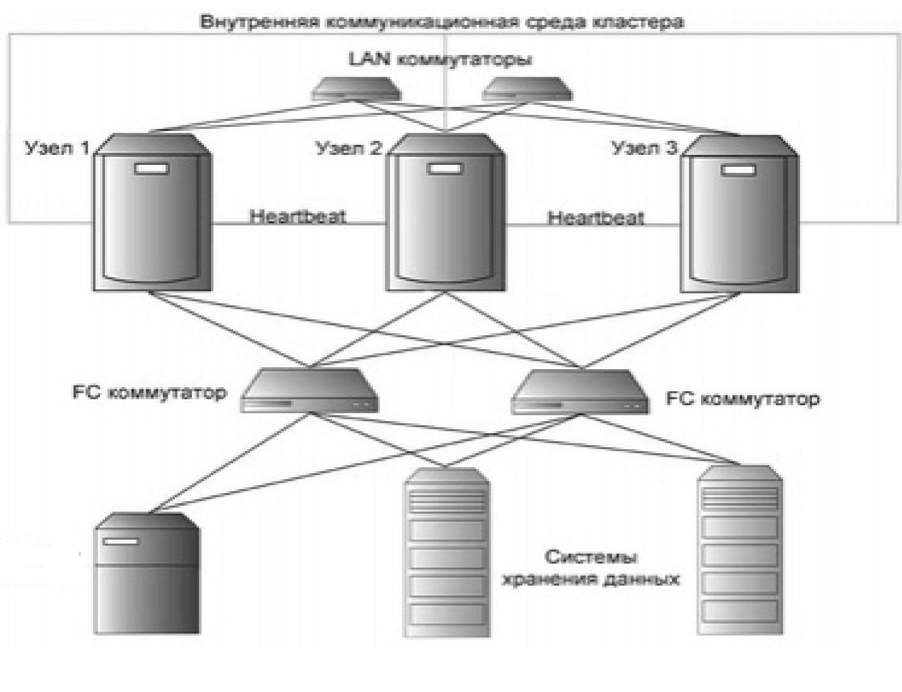

Рис. 1. Многоуровневая компьютерная система

Постановка задачи оптимизации

В качестве объекта оптимизации рассмотрим вычислительную систему, в которой выделены M уровней. Требуется найти число (кратность резервирования) узлов на каждом уровне $(m_1, m_2, ..., m_M)$, при котором обеспечивается максимум надежности системы, минимум среднего времени пребывания запросов в системе, а стоимость реализации системы не превышает выделенных средств:

$$P(m_1, m_2, ..., m_M) \to \max \text{ при } T(m_1, m_2, ..., m_M) \to \min,\ C(m_1, m_2, ..., m_M) = \sum_{i=1}^{M} m_i \le C_0.$$

Возможна также другая постановка задачи. Требуется найти число узлов на каждом уровне, при котором обеспечивается минимум стоимости реализации:

$$C(m_1, m_2, ..., m_M) \to \min \text{ при } P(m_1, m_2, ..., m_M) \ge P_0,\ T(m_1, m_2, ..., m_M) \le T_0.$$

При этом T_0, P_0, C_0 – предельно допустимые значения среднего времени ожидания, надежности и стоимости системы. При оптимизации будем считать заданными интенсивность входного потока запросов λ, показатели надежности узлов $p_1, p_2, p_3, ..., p_M$ и средние времена выполнения запросов узлами каждого уровня $v_1, v_2, v_3, ..., v_M$.

Рассматриваемая задача оптимизации является векторной, и ее сложность обусловлена тем, что одно решение (альтернатива) может превосходить другую по одним показателям и уступать по другим. В задачах принятия решений по векторному показателю главное внимание уделяется выработке решающего правила, основанного на компромиссе между значениями компонент векторного показателя. Таким образом, проблема принятия решений по векторному показателю сопряжена с поиском концептуально обоснованного критерия выбора наилучшего решения, который для отказоустойчивых систем должен характеризовать эффективность процесса деградации системы при накоплении отказов. Формальное определение обобщенного критерия в виде аддитивной или мультипликативной свертки показателей надежности и производительности не отражает возможную потерю производительности по мере накопления отказов и дает завышенный результат, что может привести к необоснованному выбору структуры системы.

Для многоуровневых отказоустойчивых систем с возможностью накопления отказов узлов различных уровней и с образованием работоспособных состояний системы с разным уровнем ее эффективности искомый комплексный показатель будем искать в виде математического ожидания эффективности (производительности) с учетом вероятностей всех работоспособных состояний в форме

$$K^*(m_1, m_2, m_3, ..., m_M, \delta) = \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \sum_{k_3=1}^{m_3} ... \sum_{k_M=1}^{m_M} E(k_1, k_2, k_3, ..., k_M) P(k_1, k_2, k_3, ..., k_M, \delta),$$

где δ – условие работоспособности состояний системы, которое может зависеть от предельно допустимого среднего времени пребывания запросов в системе $\delta(T_0)$; $E(k_1, k_2, k_3, ..., k_M) = 1/T(k_1, k_2, k_3, ..., k_M)$ характеризует эффективность состояния при исправности $k_1, k_2, k_3, ..., k_M$ узлов на соответствующих уровнях системы. Для систем с деградацией потерю эффективности относительно исходного состояния (при исправности всех $m_1, m_2, m_3, ..., m_M$ узлов системы) можно охарактеризовать выражением $E(k_1, k_2, k_3, ..., k_M) = T(m_1, m_2, m_3, ..., m_M) / T(k_1, k_2, k_3, ..., k_M)$.

Решение задач векторной оптимизации

Известны различные подходы к решению задач векторной оптимизации [1]:

– методы скалярной свертки на основе аддитивного критерия, мультипликативного

критерия, отклонения от идеала;
- методы, основанные на ранжировании критериев, в том числе метод главного критерия, лексикографическая оптимизация, метод последовательной уступки;
- методы, основанные на использовании функции полезности;
- минимаксные методы.

Перечисленные методы являются эвристическими и не исключают субъективизм лица принимающего решение (ЛПР). Субъективность методов может привести к выбору не лучшего решения. В связи с этим целесообразна разработка объективного результирующего показателя эффективности, устанавливаемого на основе анализа назначения проектируемой системы, качество которой может характеризоваться единым показателем, зависимым от частных показателей. Сложность получения такого показателя сопряжена с тем, что требуется точно представлять назначение системы как составной части надсистемы и необходимы знания зависимости обобщенного показателя от частных. Как правило, результирующий показатель должен иметь некоторый физический смысл, подтверждающий соответствие системы ее функциональному назначению с позиции надсистемы [2].

Частные показатели качества системы

Возможности системы по обработке запросов определим по среднему времени пребывания запросов в системе. В простейшем случае каждый узел представим системой массового обслуживания типа *M*/*M*/1. Считая, что каждый запрос последовательно обслуживается одним из исправных узлов на каждом уровне системы, среднее время пребывания запросов в системе оценим как $T(m_1, m_2, m_3, ...m_M) = \sum_{i=1}^{M} \frac{\text{v}_i}{1-\text{v}_i\lambda/m_i}$, где $(m_1, m_2, m_3, ..., m_M)$ – число узлов на каждом уровне.

Надежность системы оценим по вероятности сохранения работоспособности, которая зависит от формулировки условия отказа (сохранения функционирования *Y*).

Если система работоспособна, когда исправен хотя бы один узел на каждом уровне, то вероятность работоспособности системы определяется по формуле $P(m_1, m_2, m_3, \ldots, m_M) = \prod_{i=1}^{M} (1-(1-p_i)^{m_i})$, где p_i – вероятность работоспособности узла, находящегося на *i*-м уровне системы.

Если система работоспособна, когда запрос обслуживается за время, не превосходящее заданный предельный уровень (ограничение на допустимое время выполнения запросов), то надежность системы оценим как

$$P(m_1, m_2, m_3, ..., m_M) = \sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2}\sum_{k_3=1}^{m_3}...\sum_{k_M=1}^{m_M} \delta(k_1, k_2, k_3, ..., k_M) C_{m_1}^{k_1} C_{m_2}^{k_2} C_{m_3}^{k_3} ... C_{m_M}^{k_M} \times$$

$$\times p_1^{k_1} p_2^{k_2} p_3^{k_3} ... p_M^{k_M} (1-p_1)^{m_1-k_1} (1-p_2)^{m_2-k_2} (1-p_3)^{m_3-k_3} ... (1-p_M)^{m_M-k_M},$$

где $C_{m_i}^{k_i} = m_i!/k_i!(m_i-k_i)!$, а $\delta(k_1, k_2, k_3, ..., k_M) = \begin{cases} 1, & if \quad T(k_1, k_2, k_3, ..., k_M) \le T_0, \\ 0, & if \quad T(k_1, k_2, k_3, ..., k_M) > T_0. \end{cases}$

Здесь $\delta(k_1, k_2, k_3, ..., k_M)$ – условие работоспособности состояний системы при ограничении на допустимое время выполнения запросов T_0, а $T(k_1, k_2, k_3, ..., k_M)$ – среднее время пребывания запросов в системе при исправности $k_1, k_2, k_3, ..., k_M$ узлов соответствующих уровней. При особых требованиях к времени реакции (реальный масштаб времени) в качестве условия работоспособности состояний $\delta(k_1, k_2, k_3, ..., k_M)$ может использоваться вероятность непревышения времени реакции (пребывания запросов в системе) за предельное время T_0.

Сравним результаты расчета надежности с учетом ограничений на время пребывания запросов в системе. Для трехуровневой системы (M=3) будем считать $p_1 = p_2 = p_3 = 0{,}9$, $m_1 = m_2 = m_3 = 4$ шт., $v_1 = 1$ с, $v_2 = 2$ с, $v_3 = 1$ с. Зависимость надежности $P(T_0)$ от предельно допустимого среднего времени пребывания запросов в системе T_0 приведена на рис. 2, на котором кривая 1 соответствует значению интенсивности входного потока λ=1,3 1/с, а кривая 2 – λ=1,7 1/с.

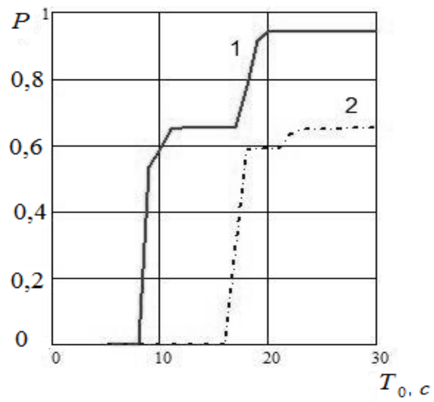

Рис. 2. Надежность системы с учетом предельного времени пребывания запросов

Формирование скалярного критерия

Скалярная свертка, используемая при решении задач векторной оптимизации, составляется формально на основе субъективных предпочтений ЛПР, и часто ее физический смысл не просматривается. В качестве скалярной свертки обычно используют мультипликативный и аддитивный критерии, определяемые для рассматриваемых систем как

$$M(m_1, m_2, ..., m_M) = \frac{P(m_1, m_2, ..., m_M)^{\alpha_1}}{T(m_1, m_2, ..., m_M)^{\alpha_2}},$$

$$A(m_1, m_2, ..., m_M) = \alpha_1 P(m_1, m_2, ..., m_M) + \alpha_2 \frac{T_0}{T(m_1, m_2, ..., m_M)},$$

где коэффициенты α_1, α_2 учитывают субъективную оценку ЛПР важности частных критериев, причем $\alpha_1 + \alpha_2 = 1$ (стоимость учитывается как ограничение).

Показатели сохранения эффективности

Эффективность каждого работоспособного состояния оценим величиной, обратной среднему времени пребывания запросов в системе при прохождении запроса через узлы каждого уровня:

$$K^*(m_1, m_2, m_3, ..., m_M) = \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \sum_{k_3=1}^{m_3} ... \sum_{k_M=1}^{m_M} \left[1/T(k_1, k_2, k_3, ..., k_M)\right] C_{m_1}^{k_1} C_{m_2}^{k_2} C_{m_3}^{k_3} ... C_{m_M}^{k_M} \times$$

$$\times p_1^{k_1} p_2^{k_2} p_3^{k_3} ... p_M^{k_M} (1-p_1)^{m_1-k_1} (1-p_2)^{m_2-k_2} (1-p_3)^{m_3-k_3} ... (1-p_M)^{m_M-k_M}.$$

Оценивая каждое работоспособное состояние отношением его эффективности относительно эффективности исходного состояния системы (когда все узлы исправны), коэффициент сохранения эффективности определим как математическое ожидание:

$$K(m_1, m_2, m_3, ..., m_M) = \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \sum_{k_3=1}^{m_3} ... \sum_{k_M=1}^{m_M} \left[\frac{T(m_1, m_2, m_3, ..., m_M)}{T(k_1, k_2, k_3, ..., k_M)} \right] C_{m_1}^{k_1} C_{m_2}^{k_2} C_{m_3}^{k_3} ... C_{m_M}^{k_M} \times$$

$$\times p_1^{k_1} p_2^{k_2} p_3^{k_3} ... p_M^{k_M} (1-p_1)^{m_1-k_1} (1-p_2)^{m_2-k_2} (1-p_3)^{m_3-k_3} ... (1-p_M)^{m_M-k_M}.$$

Коэффициент сохранения эффективности с учетом условия выполнения запроса за время, не превышающее T_0, вычислим как

$$K(m_1, m_2, m_3, ..., m_M) = \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \sum_{k_3=1}^{m_3} ... \sum_{k_M=1}^{m_M} \delta(k_1, k_2, ..., k_M) \left[\frac{T(m_1, m_2, ..., m_M)}{T(k_1, k_2, ..., k_M)} \right] C_{m_1}^{k_1} C_{m_2}^{k_2} C_{m_3}^{k_3} ... C_{m_M}^{k_M} \times$$

$$\times p_1^{k_1} p_2^{k_2} p_3^{k_3} ... p_M^{k_M} (1-p_1)^{m_1-k_1} (1-p_2)^{m_2-k_2} (1-p_3)^{m_3-k_3} ... (1-p_M)^{m_M-k_M}.$$

Зависимость коэффициента сохранения эффективности $K(T_0)$ от предельно допустимого времени пребывания запросов в системе T_0 приведена на рис 3, где кривая 1 соответствует интенсивности входного потока $\lambda = 1{,}8$ 1/с, а кривая 2 – $\lambda = 1{,}5$ 1/с.. Расчет проведен при M=3, $p_1 = p_2 = p_3 = 0{,}9$, $m_1 = m_2 = m_3 = 4$ шт., $v_1 = 1$ с, $v_2 = 2$ с, $v_3 = 1$ с.

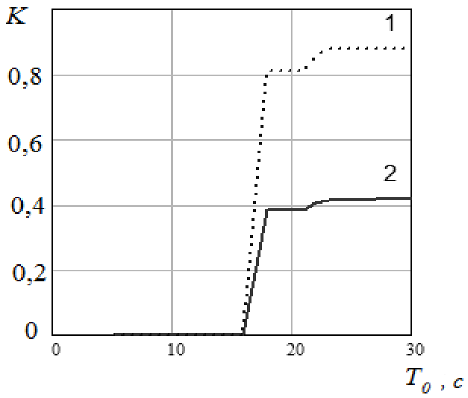

Рис. 3. Зависимость коэффициента сохранения эффективности $K(T_0)$ от предельно допустимого времени пребывания запросов в системе T_0

Результаты оптимизации

Проведем оптимизацию трехуровневой компьютерной системы по критерию максимума коэффициента сохранения эффективности, а также по максимуму аддитивного и мультипликативного критериев. Будем считать заданными C = (450, 150, 450) у.е., C_0 = 6000 у.е., V = (1, 4, 1) с, T_0 = 80 с. Результаты оптимизации сведены в таблицу. Для каждого варианта исходных данных определено распределение числа узлов на каждом уровне, которое обеспечивает соответственно максимум коэффициента сохранения эффективности, аддитивного и мультипликативного критерия при ограничении расходов на реализацию системы C_0. Оценены потенциальные потери D значения коэффициента сохранения эффективности при нахождении оптимального числа узлов на каждом уровне m_1, m_2, m_3 по максимуму аддитивного и мультипликативного критериев.

Анализ полученных данных позволяет сделать вывод о влиянии выбора критерия на результат принятия решения по построению многоуровневых компьютерных систем. При этом предпочтительно использование коэффициента сохранения эффективности, объективно отражающего требования системы по обеспечению производительности системы даже в условиях деградации производительности вследствие отказов.

λ, 1/с	Критерий	P=(0,9; 0,95; 0,9)		P=(0,8; 0,95; 0,9)	
		m_1, m_2, m_3, шт.	$K(m_1, m_2, m_3)$,D	m_1, m_2, m_3, шт.	$K(m_1, m_2, m_3)$, D
1,45	Аддитивный	3,19,4	K=0,89928 D=0,0365	4,19,3	K=0,85679811 D=0,059073881
	Мультипликативный	4,19,3	K=0,89928 D=0,0365	4,19,3	K=0,856798 D=0,0590738
	Сохранения эффективности	5,13,4	K=0,93578 D=0	5,13,4	K=0,91587199 D=0
0,45	Аддитивный	3,22,3	K=0,9821 D=0,0057	3,22,3	K=0,96700 D=0,016971
	Мультипликативный	3,22,3	K=0,9821 D=0,0057	4,19,3	K=0,978363 D=0,00561
	Сохранения эффективности	4,16,16	K=0,9879 D=0	5,13,4	K=0,983976 D=0

Таблица. Результаты оптимизации

Таким образом, показано, что при векторной оптимизации с учетом надежности, производительности и стоимости целесообразно использование обобщенного критерия на основе коэффициента сохранения эффективности. Из таблицы видна неточность решения на основе аддитивного и мультипликативного критериев (не учитывающих, что работоспособные состояния могут характеризоваться разным качеством функционирования). Следовательно, скалярные критерии мультипликативного и аддитивного типа могут использоваться при приближенном решении задачи оптимизации.

Заключение

Рассмотрены варианты формирования частных и обобщенных критериев при решении задачи векторной оптимизации многоуровневых компьютерных систем с учетом их стоимости, надежности и времени обработки запросов. Показана целесообразность применения при решении задач векторной оптимизации многоуровневых компьютерных систем критерия на основе коэффициента сохранения эффективности, учитывающего накопление отказов узлов различных уровней и наличие работоспособных состояний системы с разным уровнем эффективности (с разным средним временем пребывания запросов в системе). Оценена разница определения оптимального числа узлов на каждом уровне (кратность резервирования) при использовании в качестве целевой функции рассматриваемого критерия на основе коэффициента сохранения эффективности, а также на основе мультипликативного и аддитивного критериев.

Результаты исследований могут использоваться при проектировании и обосновании выбора многоуровневых высоконадежных компьютерных систем, в том числе кластерной архитектуры.

Литература

1. Черноруцкий И.Г. Методы принятия решений. – СПб: БХВ, 2005. – 408 с.
2. Гуткин Л.С. Оптимизация радиоэлектронных устройств по совокупности показателей качества. – М.: Сов. радио, 1975. – 368 с.

Богатырев Владимир Анатольевич – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, доктор технических наук, профессор, bva@tinuviel.ru

Богатырев Станислав Владимирович – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, аспирант, realloc@gmail.com