УДК 681.324

А.В. СКАТКОВ, Д.Ю. ВОРОНИН

Севастопольский национальный технический университет, Украина

ОБЕСПЕЧЕНИЕ ГАРАНТОСПОСОБНОСТИ РАСПРЕДЕЛЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ С ИСПОЛЬЗОВАНИЕМ ДИВЕРСНОЙ ДИСПЕТЧЕРИЗАЦИИ

Рассмотрена задача обеспечения гарантоспособности распределенной вычислительной системы. Использован диверсный подход к организации диспетчеризации в объектах критического применения. В основе предлагаемого подхода используется двукритериальная задача о назначении. Окончательное решение может быть принято как лицом, принимающим решение (ему предоставляется множество Парето-оптимальных решений), так и с использованием метода минимального отклонения от идеальной точки. Эффективность предлагаемого подхода показана при помощи рассмотрения нескольких задач анализа.

гарантоспособность, диверсность, объекты критического применения, распределенная система, диспетчеризация, задача о назначении.

Введение

Практика показала, что применение классической теории надежности к программным средствам, сетевым технологиях и web-системам не всегда достаточно эффективно. В [1] приводятся возможные причины такого положения вещей, и вводится понятие гарантоспособности. Там же поясняются базовые понятия и таксономии гарантоспособности. Известно, что гарантоспособные системы обладают рядом свойств, соблюдение которых исключает существенные материальные убытки и катастрофы различного масштаба [1]. Распределенные системы не являются исключением. От правильности их функционирования зависит безошибочная работа объектов критического применения.

Среди областей применения распределенных систем можно особо выделить следующие:

Коммерческие организации. Применение распределенных систем обусловлено довольно разветвленной структурой самих организации (наличие филиалов компании территориально разнесенных между собой). Здесь возникают задачи синхронизации данных, осуществление централизованного хранения и резервирования данных, обеспечение общего доступа к распределенным ресурсам и т.п.

Промышленные предприятия с развитым технологическим процессом. Процесс производства является многофазным: можно четко выделить все стадии изготовления изделия. Широко применяются автоматизированные системы: системы CAE (расчетов и инженерного анализа), системы CAM (проектирования технологических процессов), CAD (конструкторского проектирования), PDM (координация работы систем CAE/CAD/CAM), SCM (управление цепочками поставок), ERP (система управления предприятием), CRM (система управления взаимоотношениями с заказчиками), SCADA (диспетчеризация) и т.д. [5].

Распределенные вычисления. Являются одной из основных областей применения распределенных систем. Одним из ярких примеров является использование распределенных вычислений в процессе прогнозирования метеорологических явлений.

Задача диспетчеризации в распределенных системах в основном сконцентрирована на минимизации времени решения пакета задач. Однако в объектах критического применения эта задача приобретает особенность, связанную с сохранением гарантоспособных свойств системы. Здесь уже необходимо использовать различные методы обеспечения гарантоспособности рассматриваемого объекта.

Одним из основных методов такого рода является принцип диверсности. Известно, что для обеспечения гарантоспособности особо важных объектов метод классического резервирования ресурсов не достаточен. При классическом резервировании элементы построены по одной и той же программно-аппаратной логике и в случае воздействия на систему достаточно сильного возмущения ξ все резервные элементы выйдут из строя и это событие может привести к катастрофическим последствиям. Однако если резервные элементы будут построены с использованием различной программно-аппаратной логики, то катастрофы можно будет избежать.

Таким образом, кроме данных, необходимых для минимизации времени работы вычислительной системы необходимо каким-либо образом оценить степень критичности задачи по сравнению со всеми остальными задачами пакета для того, чтобы максимальное число диверсных узлов было занято выполнением наиболее критичных задач. Это поможет обеспечить гарантоспособность системы при наименее возможном ущербе для скорости работы объекта.

Постановка задачи

Имеется пакет $\Pi = \{З_1, З_2, \ldots, З_n\}$, который состоит из n различных задач. Задачи решаются на узлах распределенной вычислительной системы. В системе имеется n различных видов узлов $У = \{У_1, У_2, \ldots, У_n\}$.

Решение задачи представляет собой матрицу $X = \begin{bmatrix} x_{11} & x_{11} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n1} & \ldots & x_{nn} \end{bmatrix}$, где x_{ij} – булевская переменная; $x_{ij} = 1$, если j-ая задача решается на i-м узле (здесь и далее для всех x_{ij}, i – это номер узла, j – номер задачи). Иначе, $x_{ij} = 0$.

Для всех узлов системы известны времена выполнения $t_{ij} (\forall i, j \in \{1 \ldots n\})$ задач на узлах. Так как задачи решаются независимо друг от друга, то время выполнения пакета задач при определенном назначении будет равно максимальному из времен выполнения задач пакета. Таким образом, первая целевая функция $F_1 = \max_i \sum_j (t_{ij} \cdot x_{ij})$ представляет собой время выполнения пакета задач, и она подлежит минимизации.

Так как в системе сравнительно низкопроизводительные каналы связи, а задачи не могут быть разбиты на более мелкие независимые фрагменты, то справедливы ограничения:

$$\sum_j x_{ij} = 1 \ \ \forall i \ ; \ \sum_i x_{ij} = 1 \ \ \forall j \ .$$

Очевидно, что при распределении задач между узлами системы, будет играть значительную роль не только их вычислительная сложность, но также и влияние на гарантоспособность системы в целом. С тем, чтобы обеспечить гарантоспособность системы, в распределенной системе предусмотрен принцип диверсности. Он применяется к узлам распределенной системы. Необходимо так распределить задачи между узлами распределенной системы, чтобы наиболее критичные задачи решались на диверсных узлах. Имеются экспертные оценки, которые показывают желательность определенных назначений для повышения гарантоспособности распределенной системы при помощи принципа диверсности.

Эти оценки $d_{ij} (\forall i, j \in \{1,2,3,4,5\})$ в какой-то мере отражают степень критичности задачи по сравнению со всеми остальными и необходимы для того, чтобы максимальное число диверсных узлов были заняты выполнением наиболее критичных задач. Таким образом, появляется вторая целевая функция $F_2 = \sum_i \sum_j d_{ij} \cdot x_{ij}$, которая представляет собой сумму экспертных оценок желательности для всех полученных назначений и требует максимизации.

Численное решение

Пример №1. Рассмотрим случай, когда $n=5$. Времена выполнения задач на узлах $t_{ij} (\forall i, j \in \{1 \ldots 5\})$ представлены в таблице 1.

Таблица 1
Времена выполнения задач на узлах

узел \ задача	Задача 1	Задача 2	Задача 3	Задача 4	Задача 5
Узел1	4	10	6	14	3
Узел2	13	15	11	19	17
Узел3	7	14	12	8	10
Узел4	9	7	25	11	8
Узел5	15	12	17	13	16

Предположим, что четвертый ресурс представляет собой узел распределенной системы, выполненный по диверсной технологии. А третья задача управляет процессом функционирования критического объекта.

Заданы экспертные оценки

$$d_{ij} (\forall i, j \in \{1,2,3,4,5\}),$$

которые показывают желательность определенных назначений для повышения гарантоспособности распределенной системы при помощи принципа диверсности (табл. 2).

Таблица 2
Экспертные оценки, показывающие желательность определенных назначений для повышения гарантоспособности распределенной системы

узел \ задача	Задача 1	Задача 2	Задача 3	Задача 4	Задача 5
Узел 1	3	3	0	3	3
Узел 2	3	3	0	3	3
Узел 3	3	3	0	3	3
Узел 4	1	1	10	1	1
Узел5	3	3	0	3	3

Соответствующая математическая постановка задачи может быть представлена следующим образом:

$$F_1 = \max_i \sum_j (t_{ij} \cdot x_{ij}) == \max(\,(4x_{11} + 10x_{12} +$$
$$+ 6x_{13} + +14x_{14} + 3x_{15}), (13x_{21} + 15x_{22} + 11x_{23} +$$
$$+ 19x_{24} + 17x_{25}), (7x_{31} + 14x_{32}, 12x_{33} + 8x_{34} +$$
$$+ 10x_{35}), (9x_{41} + 7x_{42} + 25x_{43} + 11x_{44} + 8x_{45}),$$
$$(15x_{51} + 12x_{52} + 17x_{53} + 13x_{54} + 16x_{55})\,) \to \min_{x \in \Delta_\beta},$$

$$F_2 = \sum_i \sum_j d_{ij} \cdot x_{ij} = 3x_{11} + 3x_{12} + 0x_{13} + 3x_{14} + 3x_{15} +$$
$$+ 3x_{21} + 3x_{22} + 0x_{23} + 3x_{24} + 3x_{25} + +3x_{31} + 3x_{32} +$$
$$+ 0x_{33} + 3x_{34} + 3x_{35} + 1x_{41} + 1x_{42} + 10x_{43} + 1x_{44} +$$
$$1x_{45} + 3x_{51} + 3x_{52} + 0x_{53} + 3x_{54} + 3x_{55} \to \max_{x \in \Delta_\beta},$$

где справедливы ограничения

$$\sum_j x_{ij} = 1 \quad \forall i\,; \; \sum_i x_{ij} = 1 \quad \forall j\,.$$

Описанную математическую модель свести к двукритериальной задачи о назначении, которая относится к классу задач многокритериального булевского программирования. Ее можно решать в MS Excel методом аддитивной свертки [3]. Однако этот метод имеет недостаток, заключающийся в угнетении второй целевой функции в пользу первой. Целесообразно использовать этот метод, если оптимальность второй целевой функции малосущественна. однако, в нашем случае вторая целевая функция обеспечивает гарантоспособность распределенной системы и, исходя из этого, метод минимального отклонения от идеальной точки для этой задачи более актуален.

Таким образом, будем решать задачу в MS Excel при помощи метода минимального отклонения от идеальной точки [3].

Решение будет состоять из двух этапов. На первом этапе следует решить две однокритериальные задачи для описанных целевых функций.

При помощи мастера поиска решения было получено количественное решение задачи для целевой функции (времени). Результатом решения являются оптимальные значения переменных:

$$x_{11} = 1;\, x_{23} = 1;\, x_{35} = 1;\, x_{44} = 1;\, x_{52} = 1\,.$$

Остальные переменные равны нулю. Найденному оптимальному решению соответствует минимальное значение первой целевой функции $F_1^{опт} = 12$. Анализ полученных результатов свидетельствует о том, что первый узел следует занять первой задачей, второй – третьей, третий – пятой, четвертый – четвертый, а пятый узел необходимо занять второй задачей. При этом время выполнения пакета задач будет не меньше 12 условных единиц времени.

При помощи мастера поиска решения было получено количественное решение задачи для второй целевой функции (гарантоспособности). Результатом решения являются оптимальные значения переменных:

$$x_{11} = 1;\ x_{24} = 1;\ x_{35} = 1;\ x_{43} = 1;\ x_{52} = 1\ .$$

Остальные переменные равны нулю. Найденному оптимальному решению соответствует максимальное значение второй целевой функции $F_2^{опт} = 22$. Анализ полученных результатов свидепательствует о том, что первый узел следует занять первой задачей, второй – четвертой, третий – пятой, четвертый – третьей, а пятый узел необходимо занять второй задачей. При этом сумма экспертных оценок будет равна 22.

Для выполнения второго этапа решения двукритериальной задачи методом минимального отклонения от идеальной точки введем значения $F_1^{опт} = 12$ и $F_2^{опт} = 22$ в ячейки G3 и H3. А в ячейку I2 ввести формулу

=(G2-G3)^2+(1,5*(H2-H3))^2,

которая представляет минимальное отклонение от идеальной точки, найденной в результате решения первого этапа (1,5 – масштабирующий коэффициент, который увеличивает важность критерия, отвечающего за гарантоспособность).

При помощи мастера поиска решения было получено количественное решение двукритериальной задачи. Результатом решения являются оптимальные значения переменных:

$$x_{11} = 1;\ x_{25} = 1;\ x_{34} = 1;\ x_{43} = 1;\ x_{52} = 1\ .$$

Остальные переменные равны нулю. Найденному решению соответствует значение целевой функции времени $F_1 = 25$ и экспертной целевой функции $F_2 = F_2^{опт} = 22$.

Анализ полученных результатов свидетельствует о том, что первый узел следует занять первой задачей, второй – пятой, третий – четвертой, четвертый – третьей, а пятый – второй задачей. При этом экспертная оценка желательности определенных размещений равна 22, а время выполнения пакета задач равно 25.

Имеется конфликт между быстротой функционирования и гарантоспособностью системы. Иными словами предложенные критерии противоречивы и для обеспечения гарантоспособности системы придется решать пакет задач на 52% медленнее.

Обеспечение гарантоспособности для рассматриваемой задачи критично ко времени решения пакета задач и предлагаемый подход здесь эффективен, так как учет лишь одного критерия и игнорирование другого приведет к катастрофе. Это возможно либо за счет возникновения ошибки при решении особо важной задачи (происходит из-за игнорирования критерия, отвечающего за гарантоспособность – при решении задачи на узле, выполненном по диверсной технологии, такой ошибки бы не возникло). Либо за счет превышения критического времени отклика системы (при учете критерия, минимизирующего время решения задач в системе такой ситуации удалось бы избежать).

Пример № 2. Времена выполнения задач на узлах $t_{ij}\,(\forall i, j \in \{1 \ldots 5\})$, представлены в табл. 3.

Четвертый узел распределенной системы все также выполнен по диверсной технологии, а третья задача управляет процессом функционирования критического объекта.

Экспертные оценки $d_{ij}(\forall i, j \in \{1,2,3,4,5\})$ не изменились (см. табл. 2).

При помощи мастера поиска решения было получено количественное решение задачи для первой целевой функции (времени) и второй целевой функции (гарантоспособности). Результатом решения являются оптимальные значения переменных:

$$x_{12} = 1, x_{21} = 1, x_{35} = 1, x_{43} = 1, x_{54} = 1 .$$

Остальные переменные равны нулю. Найденному оптимальному решению соответствует минимальное значение первой целевой функции $F_1^{опт} = 12$ и второй $F_2^{опт} = 22$.

Таблица 3

Времена выполнения задач на узлах

задача / узел	Задача 1	Задача 2	Задача 3	Задача 4	Задача 5
Узел1	4	10	6	14	3
Узел2	11	15	11	19	17
Узел3	7	14	12	8	10
Узел4	9	7	3	11	8
Узел5	15	12	17	12	16

Первый узел следует занять второй задачей, второй – первой, третий – пятой, четвертый – третьей, а пятый – четвертой задачей.

При этом экспертная оценка желательности определенных размещений равна 22, а время выполнения пакета задач равно 12.

Критерии для этой задачи оказались непротиворечивыми, и вследствие этого предлагаемый подход является индифферентным по отношению к обеспечению гарантоспособности рассматриваемой системы. Опасности, что игнорирование одного из критериев может привести к катастрофе, в рассмотренном случае нет.

Парето-анализ

На рис. 1 графически представлена зависимость значений экспертной целевой функции от значений целевой функции времени. Необходимо из полученного множества точек, выделить точки, входящие в Парето-оптимальное множество. Решение многокритериальной задачи оптимизации является Парето-оптимальным, если значение любого из критериев можно улучшить лишь за счет ухудшения значений остальных критериев [4].

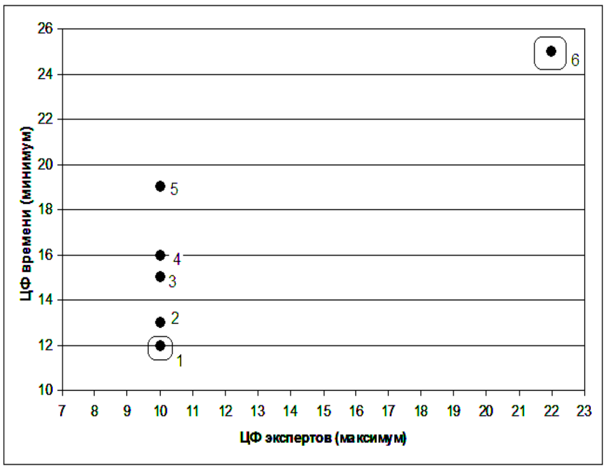


Рис. 1. Зависимость значений экспертной целевой функции от значений целевой функции времени

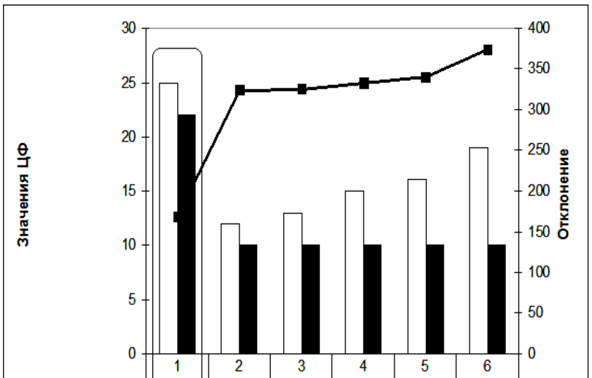
В [5] дается формальное определение Парето-оптимального множества. Пусть имеется *n* критериев, которые можно представить как пространство R^n. Элемент *x* «лучше» *y*, если

$$\forall i \in [1,n]: x_i \geq y_i , \exists k \in [1,n]: x_k > y_k .$$

Элемент $x^* \in X$ называется оптимальным по Парето, если не существует такого $x \in X$, который будет “лучше” x*.

Таким образом, согласно приведенному определению, точки под номерами «1» и «6» входят в множество Парето-оптимальных решений. Дальнейший выбор между этими двумя альтернативами делается ЛПР при помощи интуиции, сбора дополнительной информации (например, о расстановке приоритетов между противоречивыми критериями), а также при помощи методов решения многокритериальных оптимизационных задач (например, при помощи метода минимального отклонения от идеальной точки).

На рис. 2 графически представлена зависимость изменения отклонения от идеальной точки от значений целевых функций. Проанализируем эту графическую зависимость.



	1	2	3	4	5	6
ЦФ 1	25	12	13	15	16	19
ЦФ 2	22	10	10	10	10	10
Отклонение	169	324	325	333	340	373

Рис. 2. Зависимость отклонения от значений целевых функций

В первой точке отклонение минимально, однако значение целевой функции времени наихудшее, в то же время значение экспертной целевой функции наилучшее. Во второй точке значение целевой функции времени наилучшее, однако, значение экспертной целевой функции не самое наилучшее. Как было сказано выше, эти точки входят в множество Парето-оптимальных решений. Однако, если при решении руководствоваться только методом минимального отклонения от идеальной точки, то необходимо выбрать точку с минимальным отклонением от идеальной точки, то есть то решение, для которого согласованность противоречивых критериев будет максимальной. На рис. 2 это решение выделено. Отклонение от идеальной точки равно 169. Таким образом, при помощи метода минимального отклонения от идеальной точки, было получено решение задачи диспетчеризации в гарантоспособной распределенной системе.

Заключение

Предложен подход к обеспечению гарантоспособности распределенной вычислительной системы с использованием диверсной диспетчеризации. Проанализирована его эффективность.

Была оценена целесообразность предлагаемого подхода в зависимости от противоречивости критериев. Окончательное решение может быть принято как ЛПР (ему предоставляется множество Парето-оптимальных решений), так и с использованием метода минимального отклонения от идеальной точки.

Предложенный подход планируется внедрить в систему принятия решений по организации диспетчеризации в распределенных объектах критического применения.

Литература

1. Харченко В.С. Гарантоспособность и гарантоспособные системы: элементы методологии // Радиоэлектронные и компьютерные системы. – № 5 (17). – С. 7–19.

2. Норенков И.П. Основы автоматизированного проектирования. – М.: Изд-во МГТУ им. Баумана, 2002. – 336 с.

3. Леоненков А.В. Решение задач оптимизации в среде MS Excel. – Спб.: BHV-Питербург, 2005. – 704 с.

4. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач – М. Наука, 1982. – 256 с.

5. Поиск оптимальных по Парето стратегий [Электронный ресурс]. – Режим доступа: http://www.pitbear.com/category/pareto.

Поступила в редакцию 5.02.2008

Рецензент: д-р техн. наук, проф. В.Я. Копп, Севастопольский национальный технический университет, Севастополь.