МЕТОД РАЗБИЕНИЯ МНОГОМАШИННОЙ СИСТЕМЫ НА КОМПЛЕКСЫ ПРИ ОРГАНИЗАЦИИ СБОЕ- И ОТКАЗОУСТОЙЧИВЫХ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ

***И.В. Ашарина*[1], *А.В. Лобанов*[2]**

[1]*Московский государственный институт электронной техники*

[2]*ОАО «НИИ «Субмикрон», Москва*

Представлен метод выделения в многомашинной вычислительной системе или сети ЦВМ произвольной структуры заданного количества комплексов требуемой отказоустойчивости с целью организации сбое- и отказоустойчивого параллельного исполнения ряда задач, взаимообменивающихся информацией. Определены достаточные условия существования таких комплексов. Предложен новый алгебраический подход к реализации представленного метода, обеспечивающий наглядность и достоверность получаемых результатов.

Введение

Суть концепции облачных вычислений заключается в предоставлении конечным пользователям удаленного динамического доступа к услугам, вычислительным ресурсам и приложениям через Интернет. Компьютеры, осуществляющие облачные вычисления, называются вычислительным облаком. Особенностями облачных вычислений является то, что при использовании любого проприетарного ПО или чужого сервера пользователь становится беззащитным перед разработчиком этого ПО.

Другая модель облачных вычислений – т.н. частное облако – обладает многими преимуществами по сравнению с компьютерной средой на базе публичных облаков. Различие между этими моделями состоит в том, что в частном облаке управление данными и процессами осуществляется внутри организации. В этом случае отсутствуют такие проблемы, как ограничение пропускной способности сети, угрозы безопасности и необходимость нормативного соответствия, которые могли бы возникнуть при использовании публичных облаков посредством открытых сетей общего пользования.

Однако ни одно вычислительное облако не обеспечивает вычислений с требуемой степенью достоверности. Предлагаемый подход обеспечивает решение этой задачи.

1. Отказоустойчивость параллельных вычислений в сетях ЦВМ

При организации многозадачных параллельных отказоустойчивых вычислений в многомашинных вычислительных системах (МВС) и сетях ЦВМ одной из важнейших является задача достижения согласованности информации в различных ЦВМ системы в условиях возникновения допустимых неисправностей, формулируемая как проблема ***взаимного информационного согласования (ВИС)*** [1]. В данной работе рассматривается МВС произвольной структуры, отображаемой ориентированным графом (орграфом) G, ЦВМ (вершины) которой соединены двухточечными симплексными каналами связи (дугами). ***Модель неисправности ЦВМ*** – враждебная, при которой поведение неисправной ЦВМ может быть произвольным, неодинаковым по отношению к взаимодействующим с ней другим ЦВМ МВС, в том числе подобным «злонамеренному».

Комплексом k_i (i=1, 2,…, l) из заданного множества $K=\{k_1, k_2,…, k_l\}$ называется подсистема (орподграф), удовлетворяющая достаточным условиям, состоящим из трех частей [2, 3]: Ч1) наличие в k_i орподграфа H_i, гомеоморфного полному орграфу M с количеством вершин более $3m_i$, где m_i принадлежит интервалу $m_i^{\max} - m_i^{\min}$ и задает допустимое количество неисправных ЦВМ в k_i. Вершины H_i, соответствующие вершинам M, называются основными и составляют множество X_i. Остальные вершины k_i называются неосновными и составляют множество N_i; Ч2) наличие исходящих, не пересекающихся далее путей из каждой неосновной вершины k_i к не менее чем $2m_i$+1 конечным основным вершинам, причем в каждом пути имеется только одна основная вершина; Ч3) наличие входящих путей в каждую неосновную вершину k_i от не менее чем $2m_i$+1 основных вершин, причем эти пути пересекаются только в этой неосновной вершине и в каждом пути имеется только одна основная вершина, являющаяся начальной.

При сбое- и отказоустойчивом исполнении в системе l взаимообменивающихся параллельных задач необходимо наличие в системе $K=\{k_1, k_2,…, k_l\}$ комплексов, причем i-ю задачу решает k_i. В таких системах процесс достижения **системного ВИС (СВИС)** можно рассматривать в виде двух последовательных этапов: 1) **внутрикомплексное ВИС** для каждого k_i, при котором после внутрикомплексного обмена во всех исправных ЦВМ k_i вычисляется одинаковый вектор согласованных значений этого k_i, 2) **межкомплексное ВИС**, включающее межкомплексный обмен согласуемой информацией и вычисление **вектора СВИС**, одинакового в каждой исправной ЦВМ системы и содержащего вектор согласованных значений всех ЦВМ системы.

Алгоритм межкомплексного ВИС в многокомплексной МВС можно построить, если в G выделены комплексы требуемой структуры и между каждой возможной парой комплексов определена среда межкомплексных пересылок. Присутствие допустимых неисправностей в среде не должно препятствовать межкомплексному ВИС в этой паре. Ниже рассматривается только задача выделения заданного множества комплексов.

Предлагаемый метод выделения комплексов состоит в следующем. Пусть в некотором орграфе G системы необходимо последовательно выделить $K=\{k_1, k_2,…, k_l\}$. Вначале все вершины орграфа G считаются незадействованными. Выделение k_1 начинается с задания текущего значения $m_1=m_1^{\max}$. Из состава незадействованных вершин формируются всевозможные множества O_1^1, O_1^2,..., O_1^r вершин-кандидатов в основные вершины k_1 мощностью $n_1=3m_1+1$. Затем выбирается O_1^1 и среди множества всевозможных пронумерованных вариантов структур-кандидатов, определяемых O_1^1, последовательно находится вариант орподграфа H_1, удовлетворяющий достаточным условиям Ч1-Ч3 существования k_1. При положительном исходе процесс выделения k_1 прерывается и считается, что k_1 с подорграфом H_1 выделен. Из множества незадействованных вершин G исключаются вершины из k_1, и осуществляется переход к аналогичному выделению k_2 в орподграфе, порожденном в графе G оставшимися незадействованными вершинами. Иначе делается вывод о невозможности выделения k_1 для заданного O_1^1, выбирается O_1^2 и повторяется весь анализ. Если невозможно выделение k_1 также и для O_1^r, то делается вывод о невозможности выделения k_1 при заданном m_1. Тогда, если $m_1 > m_1^{\min}$, повторяются все действия для $m_1:=m_1-1$. Если же невозможно выделение k_1 при любом допустимом значении m_1, то делается вывод о невозможности выделения заданного $K=\{k_1, k_2,…, k_l\}$. Если невозможно выделение k_i (i>1), делается вывод о неудачном прерывании процесса выделения k_{i-1} и осуществляется возврат к продолжению прерванного процесса выделения k_{i-1}. Для сокращения объема вычислений перед переходом к выделению очередного k_i (i>1) выполняется проверка достаточности количества незадействованных вершин для выделения оставшихся комплексов. При недостаточности выполняется переход к продолжению прерванного выделения предыдущего комплекса. В результате применения предложенного метода осуществляется полный перебор возможных вариантов выделения $K=\{k_1, k_2,\ldots, k_l\}$ и, если имеется успешный вариант, он будет найден.

Проверку $O_i^t=\{\alpha, \gamma, \chi,..., \varepsilon\}$ на удовлетворение Ч1-Ч3 можно осуществлять следующим образом. Представим орподграф из вершины α, соединенной в орграфе *G* исходящими дугами с вершинами β, γ,..., ζ, в виде **ДНФ исходящей смежности для α в орграфе *G* (ДНФИС (α, *G*))**: $\alpha\beta\vee\alpha\gamma\vee...\vee\alpha\zeta$. В этой ДНФ не допускается использование коммутативного закона. Построим подобные ДНФИС для всех вершин орграфа *G*. Выражение ДНФИС (α, *G*) также является выражением **ДНФ путей единичной длины из α в *G* (ДНФП (1, α, *G*)).** Для формирования выражения **связки α→λ в орграфе *G* (связки (α→λ, *G*))**, отображающей всевозможные простые пути из α в λ, заменим в каждом терме из ДНФП (1, α, *G*) последний его символ на соответствующую ДНФИС и при помощи преобразования вида $(A\vee B)\cdot(C\vee D)=AC\vee AD\vee BC\vee BD$ построим **ДНФ путей двойной длины из α в *G* (ДНФП (2, α, *G*))** без применения идемпотентного и коммутативного законов, сохраняя имевшуюся последовательность символов в термах. В ДНФП (2, α, *G*) продолжим итеративно подобные подстановки и преобразования. После каждой *j*-й (*j*=1, 2,…) итерации находим в получаемой ДНФП (*j*, α, *G*) каждый терм, заканчивающийся символом λ, и переносим его (с исключением из ДНФП (*j*, α, *G*)) в виде терма в формируемое выражение **ДНФ связки α→λ внутри *G* (ДНФС (α→λ, *G*)**. Из полученной ДНФП (*j*, α, *G*) исключаем каждый терм, заканчивающийся обозначением вершины стока орграфа *G* или символом, уже имеющимся в этом терме (исключение контура), а также каждый терм, в котором появляется обозначение вершины из O_i^t, отличной от α и λ. В результате таких преобразований для любого конечного орграфа *G* за конечное число $q>0$ итераций получаемая ДНФП (*q*, α, *G*) из-за переносов и исключений становится пустой, а в ДНФС (α→λ, *G*) появляются термы, каждый из которых описывает один из путей связки, термы не повторяются, и все термы отображают всевозможные простые пути в связке (α→λ, *G*).

Логическое выражение, отображающее всевозможные **исходящие пучки (α, $O_i^t\backslash\alpha$, *G*)** путей, проходящих внутри *G* из вершины $\alpha\in O_i^t$ во все другие вершины β, χ,..., ε, из O_i^t, строится путем конъюнкции всех ДНФС (α→β, *G*), (α→γ, *G*),…, (α→ε, *G*), ее преобразования к виду ДНФ без применения идемпотентного и коммутативного законов с сохранением имеющейся последовательности символов в термах и исключением каждого терма с повторяющимися символами, отличными от символа α. В результате получается выражение **ДНФ исходящих пучков из α в вершины из $O_i^t\backslash\alpha$ внутри *G* (ДНФИП (α, $O_i^t\backslash\alpha$, *G*)**, каждый терм которого отображает совокупность путей одного исходящего пучка (α, $O_i^t\backslash\alpha$, *G*), а все термы – всевозможные такие пучки.

Метод выделения в *G* для $O_i^t=\{\alpha, \beta, \gamma,..., \varepsilon\}$ очередного k_i, такого, что для каждой вершины из O_i^t, например $\delta\in O_i^t$, существует хотя бы один исходящий пучок (δ, $O_i^t\backslash\delta$, *G*), состоит в следующем. Из всех ДНФИП (α, $O_i^t\backslash\alpha$, *G*), (β, $O_i^t\backslash\beta$, *G*), (γ, $O_i^t\backslash\gamma$, *G*),..., (ε, $O_i^t\backslash\varepsilon$, *G*) строится конъюнкция и преобразуется к виду ДНФ без применения идемпотентного и коммутативного законов с сохранением имеющейся последовательности символов в термах из ДНФИП. В результате получается **ДНФ всевозможных орподграфов в орграфе *G*, гомеоморфных полному орграфу с множеством основных**

вершин O_i^t={α, β, γ,..., ε} (ДНФОГП (O_i^t, G)), каждый терм которой отображает один такой возможный орподграф $p_i^{t,s}$ (s=1,...), а все термы – всевозможные такие орподграфы из множества $P_i^t = \{p_i^{t,1}, p_i^{t,2}, ...\}$. Каждый $p_i^{t,s}$ удовлетворяет Ч1. Для каждого $p_i^{t,s}$ все вершины из $p_i^{t,s}$, не принадлежащие O_i^t, составляют множество $N_i^{t,s}$ вершин – кандидатов в неосновные вершины k_i. Для каждой неосновной вершины должны выполняться условия Ч2 и Ч3. Обеспечение Ч2, например, для $\eta \in N_i^{t,s}$ можно достичь путем успешного построения ДНФИП (η, $O_i^{t,v}$, G) (здесь $O_i^{t,v}$ является очередным анализируемым v-м подмножеством мощностью $2m_i$+1 из всех таких пронумерованных возможных подмножеств, построенных из O_i^t), если все предыдущие $O_i^{t,1}$,…, $O_i^{t,v-1}$ оказались безуспешными. При этом возможно, что в найденных путях будут иметься внутренние вершины, не принадлежащие $N_i^{t,s}$, которые необходимо, в свою очередь, ввести в k_i (в $N_i^{t,s}$) и для каждой из них аналогично обеспечить Ч2 и Ч3.

Анализ и обеспечение Ч3 для каждой вершины, принадлежащей $N_i^{t,s}$, например η, осуществляется путем построения для очередного выбранного подмножества $O_i^{t,v}$={β, γ,..., ε} **ДНФ пучков, сходящихся в η из $O_i^{t,v}$, в G (ДНФПС (η, $O_i^{t,v}$, G))** из конъюнкции связок (β→η, G), (γ→η, G),…, (ε→η, G), каждый терм которой отображает один из искомых пучков, а все термы – всевозможные такие пучки.

Предложенный метод построения ДНФП (q, α, G) (q=1, 2,...) является, по сути, новой алгебраической формой алгоритма фронта волны [4]. При этом получаемое алгебраическое выражение можно эффективно применять для других задач анализа графов и орграфов, связанных с определением путей и их взаимоотношений.

2. Метод определения достаточной среды межкомплексной посылки при организации сбое- и отказоустойчивых параллельных вычислений в сетях ЦВМ

Второй этап достижения СВИС включает межкомплексный обмен согласуемой информацией и вычисление вектора СВИС в каждой ЦВМ системы. Рассмотрим i-й комплекс-источник, j-й комплекс-получатель и их среду межкомплексной посылки. Выделим в множество $S_{i \to j}$ каждую вершину α i-го комплекса, имеющую непустую ДНФИС (α, α∪$R_{i \to j}$∪K_j). Для каждой вершины из $R_{i \to j}$∪K_j, например β, построим ДНФИС (β, β∪$R_{i \to j}$∪K_j). По результатам внутрикомплексного ВИС в i-м комплексе-источнике все исправные вершины множества $S_{i \to j}$ должны иметь правильный вектор согласованных значений i-го комплекса. В то же время не более m_i неисправных вершин из множества $S_{i \to j}$ могут иметь произвольные значения вектора, отличные от правильного. Отсюда следует правильность следующего утверждения.

Утверждение 1. Для достижения межкомплексного ВИС необходимо $|S_{i \to j}| \geq 2m_i+1$.

Рассмотрим достаточные условия вычисления в вершине α_j определенного значения вектора W_i, совпадающего с правильным значением вектора ***согласованных значений*** V_i i-го комплекса-источника. Назовем ***вершиной 0-го ранга*** такую вершину α_j, в которой согласованное значение W_i, совпадающее с правильным V_i, может быть вычислено только по результатам посылки копий V_i согласованного значения вершин комплекса-источника в эту вершину α_j без предварительного вычисления этого согласованного значения хотя бы в одной вершине из K_j. Если в вершине α_j для вычисления правильного согласованного значения W_i необходимо использовать копию этого согласованного значения, предварительно вычисленного, например, в вершине γ_j ранга t, то эта вершина α_j приобретает ***ранг не менее t+1***.

Рассмотрим путь из граничной вершины λ_i, принадлежащий некоторому сходящемуся пучку ($U^{t,s}_{i\to j}$, α_j). Возможны следующие четыре типа такого пути: ***a*** – путь состоит только из одной дуги, образующей связь между начальной и конечной вершинами; ***b*** – все промежуточные вершины пути принадлежат орподграфу $R_{i\to j}$; ***c*** – все промежуточные вершины пути принадлежат j-му комплексу; ***d*** – часть промежуточных вершин пути принадлежит орподграфу $R_{i\to j}$, а другая часть – j-му комплексу.

Назовем орподграфом вида A (B, C, D) сходящегося пучка ($U^{t,s}_{i\to j}$, α_j) орподграф, содержащий все пути типа a (b, c, d соответственно) этого пучка и только их. Аналогично, орподграфом вида AB (AC, AD,…, ABC,…, $ABCD$) сходящегося пучка ($U^{t,s}_{i\to j}$, α_j) будет орподграф, содержащий все пути типа a и b (a и c, a и d,…, a и b и c, …, a и b и c и d соответственно) этого пучка и только их. Из простого анализа можно вывести следующие достаточные условия вычисления согласованного значения $W_i=V_i$ в вершине α_j ранга 0. **У1**) Если в орподграфе вида A сходящегося пучка ($U^{t,s}_{i\to j}$, α_j) имеются $2m_i+1$ путей, то однократной посылки копий согласованного значения i-го комплекса-источника по этим путям достаточно для вычисления в вершине α_j при помощи функции мажорирования правильного согласованного значения i-го комплекса-источника. Аналогично, достаточными условиями являются **У2**) наличие $2(m_i+m_{i\to j})+1$ путей в орподграфе вида B либо вида AB, **У3**) наличие $2(m_i+m_j)+1$ путей в орподграфе вида C либо вида AC, **У4**) наличие $2(m_i+m_j+m_{i\to j})+1$ путей в орграфе любого другого из оставшихся видов.

В дополнение к вышеприведенным типам путей, ведущим в вершину α_j из вершины $\lambda_i\in S_{i\to j}$, назовем ***путем типа e*** простой путь, ведущий из некоторой вершины γ_j в вершину α_j, внутренними вершинами которого являются только вершины, принадлежащие множеству K_j. Пусть посылка копии согласованного значения W_i, вычисленного в вершине $\gamma_j\in Z^k_{i\to j}$, в вершину α_j должна осуществляться по пути типа e. В дополнение к вышеприведенным видам орподграфов сходящегося пучка ($U^{t,s}_{i\to\alpha_j}$, α_j) назовем орподграф сходящегося пучка ($Z^k_{i\to j}$, α_j), содержащий все пути типа e этого пучка и только их, орподграфом вида E. Аналогично, орподграфом вида AE (BE,…, $ABCDE$) сходящегося пучка ($U^{t,s}_{i\to\alpha_j}$ U $Z^k_{i\to j}$, α_j) будет орподграф, содержащий все пути типа a и e (b и e,..., a и b и c и d и e соответственно) этого пучка и только их.

Достаточными условиями вычисления согласованного значения $W_i=V_i$ в вершине α_j ранга $k>0$ будут: **У5**) если в орподграфе вида E сходящегося пучка ($Z^{k,l}_{i\to j}$, α_j) имеются $2m_j+1$ путей, то однократной посылки копий согласованного значения i-го комплекса-источника по этим путям достаточно для вычисления в вершине α_j при помощи функции мажорирования правильного согласованного значения i-го комплекса-источника. Аналогично, достаточными условиями являются: **У6**) наличие $2(m_i+m_j)+1$ путей в орподграфе вида AE либо вида CE, либо вида ACE; **У7**) наличие $2(m_i+m_j+m_{i\to j})+1$ путей в орподграфе любого другого из оставшихся видов.

Орподграф, удовлетворяющий хотя бы одному из достаточных условий У1 – У7 и найденный первым в процессе поиска таких орподграфов, является орподграфом посылки α_j, и он приписывается вершине α_j в процессе поиска достаточных условий межкомплексного ВИС.

Литература

1. Генинсон Б.А., Панкова Л.А., Трантенгерц Э.А. Отказоустойчивые методы обеспечения взаимной информационной согласованности в распределенных вычислительных системах // Автоматика и телемеханика. 1989. N5. С. 3–18.
2. Ашарина И.В., Лобанов А.В., Мищенко И.Г. Взаимное информационное согласование в неполносвязных многомашинных вычислительных системах // Автоматика и телемеханика. 2003. № 5. С. 190-198.
3. Ашарина И.В., Лобанов А.В. Взаимное информационное согласование в неполносвязных гетерогенных многомашинных вычислительных системах // Автоматика и телемеханика. 2010. № 5. С. 133-146.
4. Нефедов В.Н., Осипова В.А. Курс дискретной математики: Учеб. пособие. М.: Изд-во МАИ, 1992. 264 с.