

Выделение комплексов в кластерных вычислительных системах

И.В. Ашарина

Московский государственный институт электронной техники
(Технический Университет), Москва, Зеленоград

1. Введение

Цель работы: выделение непересекающихся комплексов ЦВМ в кластерных вычислительных системах (КВС) произвольной структуры. Каждый комплекс предназначен для сбое- и отказоустойчивого решения одной из параллельно решаемых задач. Каждая из параллельных задач решается на отдельном выделенном комплексе [1]. Согласованность внутрикомплексных и межкумплексных действий гарантируется использованием механизма взаимного информационного согласования (ВИС).

Актуальность работы. При организации многозадачных параллельных вычислений в КВС одной из важнейших является задача достижения согласованности действий различных ЦВМ системы в условиях возникновения допустимых неисправностей, формулируемая как проблема достижения ВИС [2]. **Процесс ВИС** представляет собой многоаундный обмен согласуемыми значениями данных между ЦВМ с последующей их обработкой и формированием в каждой ЦВМ **вектора согласованных значений**. Существует ряд решений этой проблемы [2-6], основанных на многоаундных обменах сообщениями между ЦВМ с последующим анализом полученных сообщений и вычислением согласованных значений в каждой ЦВМ. Однако разработанные методы применяются для однокомплексных гомогенных систем [7] с дуплексными каналами межмашинной связи. В настоящей работе рассматриваются КВС произвольной структуры с каналами связи следующих видов: а) двухточечные дуплексные каналы, б) двухточечные симплексные каналы, в) каналы шинной архитектуры с широковещательным способом передачи информации (далее – широковещательный канал). Широковещательный канал присоединен к передающей по нему ЦВМ через передающее устройство сопряжения (УС-передатчик).

В качестве **модели неисправности КВС** используется наиболее общая модель враждебной неисправности ЦВМ, при которой поведение неисправной ЦВМ может быть произвольным, в том числе неодинаковым по отношению к взаимодействующим с ней другим ЦВМ КВС [3].

Комплексом называется группа ЦВМ, выполняющих копии одной и той же задачи с целью обеспечения сбое- и отказоустойчивости вычислений. **Многокомплексная КВС** – это отказоустойчивая КВС, в которой разными комплексами одновременно выполняются различные прикладные задачи, обменивающиеся данными. Комплексы пронумерованы и для каждого i -го ($i=1, 2, \dots$) комплекса задана **степень отказоустойчивости** в виде допустимого числа m_i возможных неисправных ЦВМ данного комплекса при общем числе ЦВМ в комплексе $n_i \geq 3m_i + 1$ [2].

Пусть каждая ЦВМ комплекса имеет собственное согласуемое значение. **Условия достижения ВИС в однокомплексной вычислительной системе** [2-4, 6] формулируются следующим образом: если количество неисправных ЦВМ в системе не превышает допустимого числа m , то У1) для одной и той же ЦВМ-источника согласуемой информации согласованная информация у всех исправных ЦВМ-получателей системы должна быть одинаковой; У2) для исправной ЦВМ-источника согласуемой информации согласованная информация во всех исправных ЦВМ-получателях системы должна быть равна согласуемой информации этой ЦВМ-источника. При этом число ЦВМ в вычислительной системе, участвующих в процессе ВИС, в условиях возникновения враждебных неисправностей должно быть не менее $n=3m+1$. В [2-4] основным требованием к вычислительной системе является требование полносвязности ее графовой модели. В

[6] проведен анализ условий достижения ВИС в неполносвязных одноконтактных вычислительных системах с дуплексными каналами межмашинных связей и доказано, что для достижения ВИС в таких системах достаточно, чтобы графовая модель системы включала подграф, гомеоморфный полносвязному с числом основных вершин не менее $3m+1$. Там же представлен и обобщенный алгоритм ВИС для неполносвязных вычислительных систем (алгоритм АНС-1), который является основой для построения детализированного алгоритма ВИС для конкретного графа системы.

Рассмотрим **орграфовую модель** рассматриваемой многоконтактной КВС. **Передача информации** из ЦВМ-передатчика s в соседнюю ЦВМ-приемник r осуществляется по симплексному каналу связи между ними или по псевдосимплексным каналам через УС-передатчик ЦВМ s [7]. **Посылка информации** из ЦВМ-источника s в ЦВМ-получатель r по некоторому пути посылки выполняется посредством последовательности передач информации между соседними ЦВМ в этом пути. Каждому такому возможному пути посылки информации соответствует некоторый путь в орграфе КВС и наоборот.

Комплексом-источником данных называется комплекс, передающий свое согласованное значение (вектор согласованных значений) другому комплексу, который называется **комплексом-получателем** данных и воспринимает полученное значение в качестве собственного согласуемого значения. Это согласуемое значение комплекс-получатель должен сделать собственным согласованным значением. **Средой межкомплексной посылки информации** из комплекса-источника i в комплекс-получатель j (**средой $i \rightarrow j$**) называется [7] совокупность всех имеющихся в графе КВС путей таких, что каждый из путей начинается в вершине комплекса i , заканчивается в вершине комплекса j , ни одна другая вершина в пути не принадлежит этим комплексам и ни одна вершина не повторяется в одном и том же пути. **Граничными ЦВМ i -го комплекса в среде $i \rightarrow j$** называются все ЦВМ этого комплекса, принадлежащие данной среде. Граничные ЦВМ комплекса обеспечивают выполнение процесса **межкомплексного обмена** (МКО).

Условия достижения ВИС в многоконтактной КВС формулируются следующим образом: если количество неисправных ЦВМ в j -м комплексе-получателе не превышает допустимого числа m_j , то для одной и той же ЦВМ-источника согласуемого значения из i -го комплекса-источника согласованное значение у всех исправных ЦВМ комплекса-получателя должно совпадать с согласованным значением этой ЦВМ в комплексе-источнике.

В данной работе решается следующая задача. Дана произвольная структура гетерогенной КВС. Необходимо в этой структуре выделить N пронумерованных неп пересекающихся комплексов. Для каждого i -го комплекса задан диапазон степеней отказоустойчивости от $m_i^{\max}$ до $m_i^{\min}$, набор сочетаний значений отказоустойчивости всех комплексов образует **вектор степеней отказоустойчивости**, в соответствии с которыми должны быть выделены комплексы. Все возможные вектора степеней отказоустойчивости образуют упорядоченное множество, в котором осуществляется полный перебор сочетаний значений степеней отказоустойчивости всех комплексов, упорядоченных по убыванию, начиная с комплекса, имеющего максимальный номер.

2. Использование алгоритма Тэрри для выделения комплексов в КВС

Для выделения подорграфа, гомеоморфного полносвязному, воспользуемся алгоритмом Тэрри [9] поиска пути между двумя заданными вершинами v и w .

Дополнения и ограничения к алгоритму Тэрри для поиска пути между основными ЦВМ КВС:

- 1) граф, являющийся моделью КВС, является ориентированным,
- 2) искомые пути между основными вершинами [6] орграфа должны быть непересекающимися,
- 3) пути между основными вершинами могут проходить только через неосновные [6] вершины,
- 4) если есть дуга между основными вершинами, другие пути искать не следует.

Модификация алгоритма Тэрри для ориентированного подграфа (рис.1)

1. При проходе по произвольной дуге, всякий раз отмечать эту дугу.
2. Из некоторой вершины v' , всегда следовать только по той дуге, которая не была пройдена.
3. Если из некоторой вершины v'' , следующей за вершиной v' и отличной от w , нет непройденных исходящих дуг, то этот путь не ведет к вершине w . Тогда происходит возврат к вершине v' и переход к шагу 2.
4. Пометить полученный путь таким образом, чтобы исключить повторное прохождение принадлежащих ему неосновных вершин. Конец алгоритма.
- 5.

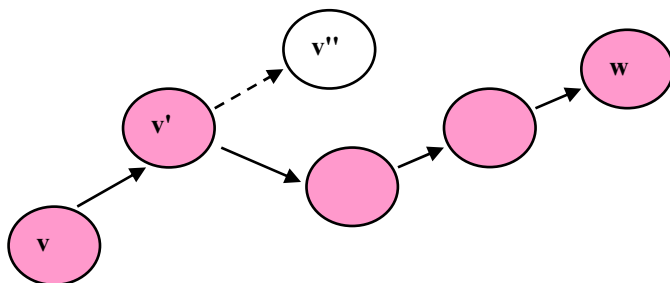


Рисунок 1. Иллюстрация работы модифицированного алгоритма Тэрри

Достоверность работы алгоритма подтверждается полным перебором путей в случае необнаружения искомого пути, характерным для алгоритма Тэрри в его классическом представлении.

Приведенный метод использования модифицированного алгоритма Тэрри поиска путей в орграфе дает возможность выделения комплексов в кластерных гетерогенных системах при заданных значениях степеней отказоустойчивости комплексов.

Для полного перебора возможных вариантов разбиения, т.е. для выбора основных вершин каждого из комплексов воспользуемся следующим подходом. Комплекс представляет собой ориентированный подграф, гомеоморфный полносвязному, поэтому не менее n_i его основных вершин должны иметь количество входящих дуг не менее чем $n_i - 1$ и количество исходящих дуг не менее чем $n_i - 1$, здесь n_i – число вершин в i -м комплексе. Последовательность действий по выделению комплексов такова. Определяем количество вершин с числом входящих дуг и числом исходящих дуг, достаточным для того, чтобы считать эти вершины основными. В соответствии с необходимой степенью отказоустойчивости, выбираем множество основных вершин первого комплекса и пытаемся сформировать на его основе комплекс, для чего, пользуясь модифицированным алгоритмом Тэрри, находим непересекающиеся пути от каждой его основной вершины к каждой другой его основной вершине. Если это не удалось, выбираем другое множество основных вершин для этого комплекса. Поступаем так до тех пор, пока не будет сформирован первый комплекс или пока не выполним полный перебор множеств, соответствующих первому комплексу.

Если первый комплекс сформирован, подобным же образом формируем второй комплекс и т.д. При невозможности сформировать второй комплекс, возвращаемся на шаг назад и формируем первый комплекс на основе другого множества основных вершин. Повторяем перечисленные действия пока не будут сформированы все комплексы или пока не будет доказана (путем полного перебора вариантов) невозможность разбиения данной КВС на указанное количество комплексов для данного вектора степеней отказоустойчивости.

Если разбиение КВС для данного вектора степеней отказоустойчивости выполнить не удалось, переходим к следующему вектору степеней отказоустойчивости из упорядоченного множества (с меньшими значениями степеней отказоустойчивости). Повторяем все перечисленные действия до тех пор, пока не будет выполнено разбиение КВС на комплексы или пока путем полного перебора всех возможных вариантов не будет доказана невозможность такого разбиения.

Пример. Использование модифицированного алгоритма Тэрри для двухкомплексной КВС.

Рассмотрим КВС, орграф которой представлен на рис.2. Отказоустойчивость каждого из комплексов принимаем равной $m_1=m_2=1$, поэтому в каждом из комплексов достаточно иметь по 4 вершины, соответствующие основным ЦВМ.

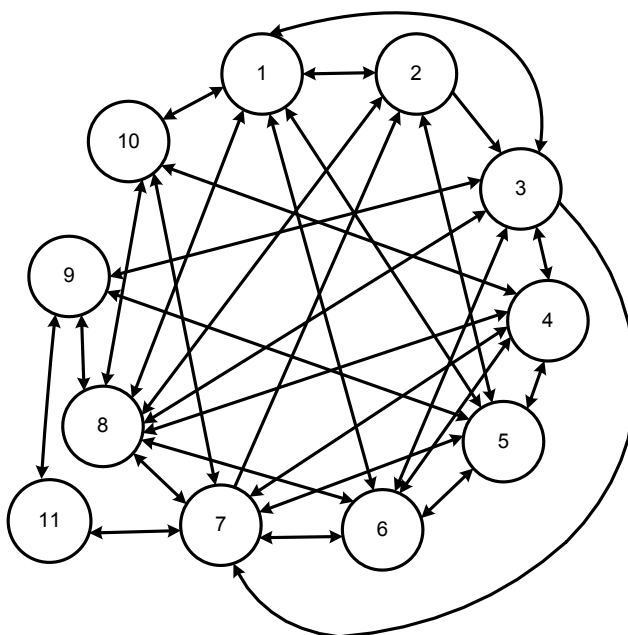


Рисунок 2. Пример двухкомплексной МВС

В качестве основных вершин первого комплекса выбираем вершины 3, 7, 8, 9. В соответствии с модифицированным алгоритмом Тэрри находим пути от каждой основной вершины комплекса 1 к каждой другой основной вершине этого комплекса: $3 \rightarrow 7$, $3 \rightarrow 8$, $3 \rightarrow 9$, $7 \rightarrow 2 \rightarrow 3$, $7 \rightarrow 8$, $7 \rightarrow 11 \rightarrow 9$, $8 \rightarrow 3$, $8 \rightarrow 7$, $8 \rightarrow 9$, $9 \rightarrow 3$, $9 \rightarrow 8$, $9 \rightarrow 11 \rightarrow 7$.

В качестве основных вершин второго комплекса выбираем вершины 4, 5, 6, 10. В этом случае нет возможности найти непересекающиеся пути от каждой основной вершины комплекса 2 к каждой другой основной вершине этого комплекса.

Выбираем другое множество основных вершин второго комплекса: 1, 4, 5, 6. В этом случае пути от каждой основной вершины комплекса 2 к каждой другой основной

вершине этого комплекса будут следующими: $1 \rightarrow 10 \rightarrow 4$, $1 \rightarrow 5$, $1 \rightarrow 6$, $4 \rightarrow 10 \rightarrow 1$, $4 \rightarrow 5$, $4 \rightarrow 6$, $5 \rightarrow 1$, $5 \rightarrow 4$, $5 \rightarrow 6$, $6 \rightarrow 1$, $6 \rightarrow 4$, $6 \rightarrow 5$.

Таким образом, выделили в КВС два комплекса.

В качестве дальнейших направлений работы предлагается: 1) анализ свойств среды межкомплексной посылки, обеспечивающих возможность вычисления согласованного значения в комплексе-получателе, 2) способ выделения среды межкомплексной посылки в КВС, 3) построение алгоритма ВИС для многокомплексных КВС.

Литература

1. Лобанов А.В. Проблемы организации достоверных вычислений в кластерных системах и сетях ЦВМ. Данный сборник.
2. Генинсон Б.А., Панкова Л.А., Трахтенгерц Э.А. Отказоустойчивые методы обеспечения взаимной информационной согласованности в распределенных вычислительных системах // *АиТ*. 1989. № 5. С. 3-18.
3. Pease M., Shostak R., Lamport L. Reaching agreement in the presence of faults // *J. Ass. Comput. Mach.* 1980. V. 27. № 2. P. 228-237.
4. Lamport L., Shostak R., Pease M. The byzantine generals problem // *ACM Trans. Progr. Lang. and Syst.* 1982. V. 4. № 3. P. 382-401.
5. Barborak M., Malek M. The consensus problem in fault-tolerant computing // *ACM Computing Surveys*. June 1993. V. 25. № 2. P. 171-220.
6. Ашарина И.В., Лобанов А.В., Мищенко И.Г. Взаимное информационное согласование в неполносвязных многомашинных вычислительных системах // *АиТ*. 2003. № 5. С. 190-199.
7. Лобанов А.В. Модели замкнутых многомашинных вычислительных систем со сбоем и отказоустойчивостью на основе репликации задач в условиях возникновения враждебных неисправностей // *АиТ*. 2009. №2. С. 171-189.
8. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. – М.: Мир, 1979.-536 с.
9. Нефедов В.Н., Осипова В.А. Курс дискретной математики: Учеб. пособие. – М.: Изд-во МАИ, 1992.-264 с.