На правах рукописи

Во Минь Тунг

Исследование кластерных вычислительных систем и разработка моделей назначения фрагментов параллельных программ

Специальность: 05.13.15 – Вычислительные машины, комплексы и компьютерные сети

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Москва – 2010

Работа выполнена на кафедре Вычислительных машин, систем и сетей Московского энергетического института (технического университета)

Научный руководитель: кандидат технических наук
профессор
Ладыгин Игорь Иванович

Официальные оппоненты: доктор технических наук
профессор
Кутепов Виталий Павлович

кандидат физико–математических наук
доцент
Ковтушенко Александр Петрович

Ведущая организация: ОАО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева»

Защита состоится 24 июня 2010 г. в 16 час. 00 мин. на заседании диссертационного совета Д 212.157.16 при Московском энергетическом институте (техническом университете) по адресу 111250, г. Москва, ул. Красноказарменная, д. 13, ауд. М-510

С диссертацией можно ознакомиться в библиотеке Московского энергетического института (технического университета).

Отзывы в двух экземплярах, заверенные печатью, просьба направлять по адресу: 111250, г. Москва, ул. Красноказарменная, д. 14, Ученый Совет МЭИ (ТУ).

Автореферат разослан «21» мая 2010 г.

Ученый секретарь
диссертационного совета Д 212.157.16
кандидат технических наук
доцент С.А. Чернов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

В настоящее время кластерные вычислительные системы (КВС) являются самыми распространенными из широко используемых суперкомпьютерных вычислительных систем в мире. Тем не менее, эффективное применение КВС на практике представляет собой не простую задачу. Обусловлено это множеством причин. Одной из причин, которая влияет на эффективность применения КВС, является не достаточно обоснованное решение задачи назначения фрагментов параллельных программ (ФПП) на КВС. При неудачном назначении, время выполнения задачи намного возрастает за счет накладных расходов. Под накладными расходами в работе понимаются те временны*е* затраты, которые определяются факторами, замедляющими выполнение задачи. Это непроизводительное время, связанное с обменом блоками данных между оперативной и дисковой памятью, время на разрешение конфликтов, которые возникают при учете приоритетов на обращения к памяти и передаче данных между коммутационными сетями (КС).

Актуальность работы. Существующие методы решения задачи назначения ФПП на вычислители КВС комплексно не учитывают вышеперечисленные факторы. В ряде работ предлагаются эвристические алгоритмы, учитывающие только номинальную производительность вычислителей, а иногда и время передачи данных по коммутационным сетям разных уровней. В настоящей работе предлагается в качестве дополнительных характеристик в процессе статического назначения ФПП учитывать приведенные выше факторы, представленные коэффициентами накладных расходов, что позволит повысить эффективность выполнения параллельных программ (ПП) на КВС. Поэтому разработка методики назначения ФПП на вычислители КВС, обеспечивающей учет таких важных характеристик конкретных КВС, как состояние памяти вычислительного узла в момент обработки задачи и неоднородность коммуникационной среды, выраженных в виде коэффициентов накладных расходов, является актуальной задачей.

Объектом исследования в работе являются кластерные вычислительные системы на примере кластеров МЭИ и Ханойского Технологического Университета (ХТУ), **предметом исследования** — особенности организации памяти и КС КВС, учет которых позволит сократить время выполнения параллельных программ на КВС за счет эффективного назначения ФПП на выделенные вычислительные ресурсы.

Цель работы и задачи исследования.

Цель диссертационной работы заключается в исследовании КВС для получения численных значений коэффициентов накладных расходов, разработка математической модели и методики статического назначения ФПП, учитывающих как характеристики задач, КВС, так и значения данных коэффициентов.

Поставленная цель определяет следующие основные задачи исследования.

1. Анализ современных КВС с точки зрения оценки влияния особенностей

иерархически-неоднородной организации памяти и КС КВС на время выполнения прикладных задач, а также анализ различных методов и моделей распределения фрагментов параллельных программ по узлам КВС.

2. Разработка математической модели оптимизации назначения ФПП на выделенные ресурсы КВС с учетом характеристик КВС, задач и накладных расходов, возникающих при параллельных вычислениях.

3. Проведение тестирования указанных выше реальных КВС с целью получения экспериментальных данных, представленных в виде коэффициентов накладных расходов.

4. Разработка имитационной модели процесса выполнения задач на КВС и алгоритма статического назначения ФПП на вычислители КВС, основанного на аппарате теории генетических алгоритмов, учитывающего особенности организация памяти и КС, для КВС МЭИ и ХТУ.

Научная новизна результатов, полученных в диссертации.

1. Определенны особенности иерархически-неоднородной организации памяти и коммуникационной среды КВС, а также виды накладных расходов, возникающих на разных уровнях иерархии, влияющих на время выполнения параллельных программ.

2. Разработана математическая модель задачи оптимизации обобщенного вида, выраженная в аналитической форме, которая отражает зависимость времени выполнения ПП от варианта назначения ФПП на выделенные вычислители КВС.

3. Проведена адаптация генетического алгоритма для задачи эффективного назначения ФПП на вычислители КВС, построена имитационная модель процесса выполнения ПП на КВС и разработана методика статического назначения ФПП на выделенные вычислители КВС.

Практическая значимость работы заключается в создании программных средств, реализующих предложенную автором методику, которые могут быть использованы при разработке параллельных программ и планировании их выполнения на выделенных ресурсах КВС для повышения эффективности использования конкретных систем. Получены реальные значения времени обращения к памяти на разных уровнях ее иерархии и пропускной способности КС КВС МЭИ и ХТУ. Определены особенности конфликтов, возникающих при одновременном обращении к памяти на разных уровнях КС, которые приводят к увеличению времени решения задачи.

Внедрение результатов работы.

Результаты диссертационной работы используются в учебном процессе кафедры ВМСиС МЭИ (ТУ) при проведении лекционных и лабораторных занятий по курсу «Вычислительные системы». Они также применяются в учебном процессе Ханойского Технологического Университета и Института Информационных Технологий при Вьетнамской Академии Наук и Технологий для проведения лекционных и лабораторных занятий по курсу «Вычислительные системы».

Достоверность результатов работы подтверждена экспериментальным исследованием решения различных задач на кластерах МЭИ, МГУ и ХТУ.

Апробация работы. Основные положения и результаты диссертации докладывались и обсуждались на 15-й Международной научно-технической конференции «Информационные средства и технологии», г. Москва, 2008 г., а также на 16-й Международной научно-технической конференции студентов и аспирантов, г. Москва, 2010 г.

Публикации. Основные результаты, полученные при выполнении диссертационной работы, опубликованы в 3 печатных работах.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка используемых источников и 9 приложений. Диссертация содержит 218 страниц машинного текста, включая приложения.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснованы актуальность и научная новизна работы, сформулированы цель работы, практическая значимость, основные задачи, решаемые в диссертации, указана область применения разрабатываемой методики решения задачи назначения.

В **первой главе** приводится классификация параллельных вычислительных систем. Проводится обзор наиболее мощных суперкомпьютерных систем СНГ, из которого следует, что КВС являются самыми популярными на данный момент времени. Даются основные определения, связанные с современными КВС. Рассмотрены принципы построения КВС на разных аппаратных платформах, для которых характерны иерархическая организация памяти и различные пропускные способности каналов связи между вычислительными ресурсами — вычислительными узлами (ВУ), процессорами и их ядрами. Делается вывод, что современные КВС с одной стороны являются мультиархитектурными, строящимися по иерархическому принципу организации, а с другой стороны, по типу применяемых аппаратных технологий и методов назначения ФПП, являются уникальными системами обработки данных. Следовательно, при выборе тех или иных технологий построения КВС может существенно варьироваться производительность системы в целом. Поэтому, получая удовлетворительный результат решения задачи на одной КВС, можно получить не удовлетворительный результат, перенося ее решение на другую КВС. На примере КВС МЭИ и ХТУ проведен анализ иерархически-неоднородной организации памяти и КС.

Время выполнения ПП на КВС существенно зависит от того, насколько эффективно назначены ФПП на вычислители КВС. Под эффективным назначением понимается такое распределение ФПП по вычислителям КВС, при котором достигается минимум времени выполнения задачи за счет учета особенностей решаемой задачи, организации иерархически-неоднородной памяти, КС и накладных расходов, возникающих на разных уровнях иерархии при выполнении ПП.

Задача назначения относится по своей природе к сложным комбинаторным задачам составления расписаний и ее решение в настоящее время получено лишь для узкого класса параллельных алгоритмов (ПА) или для конкретных вычислительных систем (ВС). Применимость известных методов и алгоритмов назначения ФПП для современных КВС осложнена тем, что современные КВС имеют особенности, которые в них не учитываются. В некоторых работах предлагаются эвристические алгоритмы, учитывающие только номинальную производительность вычислителей, а иногда и время передачи данных по коммутационным сетям разных уровней. В некоторых работах игнорируют объемы как обрабатываемых, так и передаваемых данных между ФПП. Существующие в свободно распространяемых средствах системы планирования заданий, где задача назначения решается с помощью набора правил и алгоритмов, не учитывают характеристики или структуру задач решаемого класса, неоднородность КС и памяти, накладные расходы, существующих на разных уровнях иерархии системы. Однако, учет вышеперечисленных параметров необходим, так как это может отразиться на времени выполнения ПП. Поэтому предлагается в качестве дополнительных характеристик в процессе статического назначения фрагментов ПП учитывать приведенные выше факторы.

В разделе 1.9 представлена постановка решаемой в диссертации задачи. Данная задача состоит в исследовании КВС и разработке алгоритма статического назначения ФПП на вычислители КВС с заданными характеристиками на основе генетического алгоритма. Формально данная задача определяется как поиск такого варианта назначения ФПП на вычислители КВС, при котором достигается минимальное значение времени выполнения ПП при заданных ресурсах КВС.

Для решения поставленной задачи предложено формальное описание КВС и выполняемых на них задач. В диссертации рассмотрены две моделей КВС, соответствующие различным, применяемым в настоящее время аппаратным платформам. В основной модели КВС представляется в виде дерева (рис. 1), состоящего из SN уровней $S = \{s_1, s_2, \ldots, s_{SN}\}$. Каждый уровень дерева ассоциирован с определенными вычислительными ресурсами, которые объединены КС соответствующего уровня иерархии. В нашем случае (рис. 1), первый уровень представляет саму КВС, которая состоит из n вычислительных узлов, каждый узел состоит из m процессоров, а процессоры состоят из k ядер. Таким образом, КВС состоит из $NK = n \times m \times k$ ядер (заметим, что в данной работе термины «ядро» и «вычислитель» используются как синонимы). Для каждого уровня $s_r \in S$ КС известен показатель производительности B_{sr} ($[B_{sr}]$ = байт/сек).

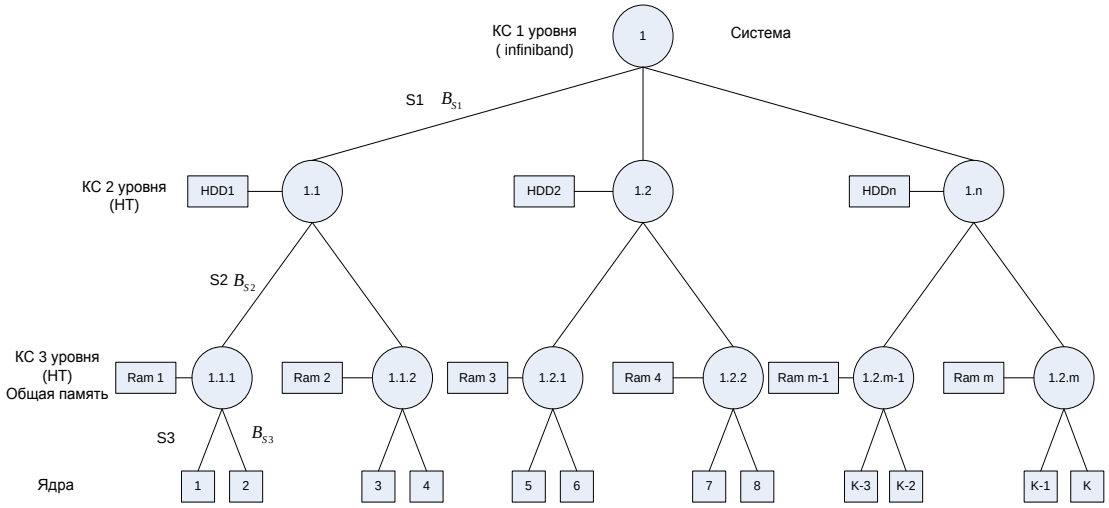

Рис. 1. Модельное представление КВС с иерархической организацией памяти и КС

Параллельный алгоритм в диссертации представляется в виде ациклического графа. Для обозначения графа применяется стандартная символика $G = <A, E>$, где $A = \{a_1, \ldots, a_i, \ldots, a_N\}$ – конечное множество вершин, а E – отношение инцидентности между парами, отображающее информационно-логические связи между вершинами. Вершины графа ПА соответствуют фрагментам ПП, а дуги – связям по данным между вершинами. Вершины взвешены временем выполнения t_i на вычислителе с заданным быстродействием и емкостью памяти V_i, требуемой для хранения как командной последовательности i-ой вершины, так и ее данных. Граф алгоритма не имеет циклов и условий, он имеет только одну начальную и только одну конечную вершину, соответствующие началу и концу выполнения ПА.

Назначение вершин графа алгоритма на вычислители КВС, определяется соответствующими значениями матрицы назначения X, $X = \{x_{id} \mid i \in [1, N], d \in [1, K]\}$; $x_{id} = 1$, если i вершина назначена на d вычислитель, и $x_{id} = 0$ в противном случае. Необходимо назначить фрагменты ПП на выбранные K вычислителей из NK вычислителей системы таким образом, чтобы время решения было минимальным (близким к минимальному) с учетом характеристик задачи, характеристик КВС и вышеперечисленных факторов, которые влияют на время выполнения ПП. Такой вариант назначения в работе называется эффективным назначением. Каждая задача, решаемая на КВС, может характеризоваться максимальным теоретическим временем выполнения и минимальным теоретическим временем выполнения. Считается, что минимальное теоретическое время выполнения $T_{KP(теорим)}$ определяется как сумма времен выполнения вершин графа, принадлежащих его критическому пути, без учетов обменов между вершинами. Максимальное теоретическое время выполнения $T_{посл}$ определяется временем выполнения задачи на одном

вычислителе. Следовательно, справедливо следующее соотношение: $T_{КР(теорит)} < T\ (X) < T_{посл}$, где *T(X)* время решения ПП при варианте назначения *X*.

Таким образом, решение поставленной в диссертации задачи заключается в разработке алгоритма поиска варианта эффективного назначения, при заданных характеристиках задачи и КВС, т.е. такого варианта назначения, при котором время выполнения задачи является минимальным среди всех рассматриваемых вариантов назначениях *X*.

Вторая глава посвящена экспериментальному исследованию характеристик КВС. Главное внимание уделено детальному исследованию влияния отмеченных выше факторов на время выполнения ПП с помощью разработанных синтетических тестов. Предложено, в качестве дополнительных характеристик в процессе статического назначения фрагментов ПП, учитывать указанные факторы, которые могут быть отображены в виде коэффициентов накладных расходов.

Введены следующие коэффициенты накладных расходов, определяемых для каждого уровня иерархии при обращении к памяти $\varepsilon_1 = \frac{\Delta t_1}{T_1}$ и $\varepsilon_2 = \frac{\Delta t_2}{T_2}$ передачи данных. Здесь Δt_1 - временные затраты, определяемые накладными расходами из-за конфликтов при обращении к используемому уровню памяти, T_1 - время обращения к используемому уровню памяти без конфликтов. Δt_2 - временные затраты, определяемые накладными расходами из-за конфликтов при передаче данных на используемый уровень КС, T_2 - время передачи данных без конфликтов на используемый уровень КС. Описаны разработанные программные средства тестирования КВС, служащие для определения времени передачи данных и обращения к памяти, а так же вычисления накладных расходов возникающих на разных уровнях иерархии. В разделе 2.2 приведены результаты экспериментального исследования времени обращения к памяти, по которым определяются значения введенных автором коэффициентов замедления времени обращения к памяти на разных уровнях иерархии $\partial_1^{'}, \partial_1^{''}, \partial_1^{'''}$ (КЭШ, ОЗУ, внешний носитель). Так же определены значения коэффициентов накладных расходов $\varepsilon_1^{'}, \varepsilon_{1,}^{''}\varepsilon_1^{'''}$, возникающих при конфликтах одновременного обращения к памяти на разных уровнях иерархии для КВС МЭИ и ХТУ. В качестве примера, в таблицах 1 и 2 представлены значения коэффициентов накладных расходов для КВС МЭИ и ХТУ, построенных на разных аппаратных платформах (вычислительные узлы КВС МЭИ построены на основе двух двуядерных процессоров *AMD Opteron* 254, а КВС ХТУ построены на двух двуядерных процессорах *Xeon E*5410), при обращении к памяти, для разных значений объема обрабатываемых данных, числа вычислителей и их местоположения в КВС. Выявлено, что если в одном ВУ выполняются две пользовательские ПП от разных пользователей, время выполнения ПП увеличивается. Автор предполагает, что данный фактор связан с распределением ресурсов операционной системой между пользователями при мультипрограммном режиме.

Значение коэффициентов накладных расходов ε_1 КВС МЭИ Таблица 1.

Кол-во ядер	Число ВУ х число вычислителей	от 50КБ-5000Кб				от 2Мб-3800Мб				4,6 Гб до 6 Гб			
		номера вычислителей				номера вычислителей				номера вычислителей			
		1	2	3	4	1	2	3	4	1	2	3	4
2	1x2	0	0,005	-	-	0,03	0,026	-	-	-	-	-	-
	2x1	0,01	0,006	-	-	0,1	0,008	-	-	-	-	-	-
3	1x3	0,01	0,003	0,008	-	0,046	0,048	0,01	-	-	-	-	-
	3x1	0,008	0,002	0,01	-	0,003	0,11	0,008	-	-	-	-	-
4	1x4	0,01	0,002	0,01	0,003	0,058	0,075	0,06	0,065	0,05	0,01	0,06	0,01
	2x2	0,003	0,008	0,001	0,002	0,06	0,075	0,1	0,078	-	-	-	-
	4x1	0	0	0,01	0,005	0,001	0,01	0,005	0,007	-	-	-	-

Значение коэффициентов накладных расходов ε_1 КВС ХТУ Таблица 2.

Кол-во ядер	Число ВУ х число вычислителей	от 2000КБ-5000Кб				от 12Мб-1900Мб				4,6 Гб до 6 Гб			
		номера вычислителей				номера вычислителей				номера вычислителей			
		1	2	3	4	1	2	3	4	1	2	3	4
2	1x2	0,027	0,006	-	-	0,3	0,35	-	-	-	-	-	-
	2x1	0,1	0,1	-	-	0,005	0,01	-	-	-	-	-	-
3	1x3	0,18	0,02	0,18	-	0,37	0,18	0,4	-	-	-	-	-
	3x1	0,02	0,005	0,01	-	0,012	0,0006	0,01	-	-	-	-	-
4	1x4	0,19	0,16	0,15	0,15	0,098	0,67	0,72	0,089	0,05	0,2	1,04	1,01
	2x2	0,04	0,01	0,033	0,006	0,37	0,24	0,23	0,21	-	-	-	-

По полученным в экспериментах данным определено, что значение коэффициентов $\partial_1^{'}, \partial_1^{''}, \partial_1^{'''}$ для КВС ХТУ меньше, чем для КВС МЭИ. Но как только возникают конфликты при одновременном обращении к памяти всех четырех вычислителей, время обращения к памяти для ВУ КВС ХТУ резко возрастает, и превышает время обращения к памяти ВУ КВС МЭИ. Очевидно, что влияние особенности организации памяти в ВУ может привести к увеличению времени выполнения ПП. В разделе 2.3 приведены результаты экспериментального исследования времени передачи данных на разных уровнях иерархии. Определены реальные времена передачи данных, по которым были определены коэффициенты замедления времени передачи данных на разных уровнях иерархиях $\partial_2^{'}, \partial_2^{''}, \partial_2^{'''}$ (между вычислителями, находящимися в одном процессоре, между вычислителями находящимися в разных процессорах, но в одном ВУ, между вычислителями в разных ВУ). Для КВС МЭИ $\partial_2^{'} = \partial_2^{''} = 1$, $\partial_2^{'''} = 2$; Для КВС ХТУ $\partial_2^{'} = \partial_2^{''} = 1$, $\partial_2^{'''} = 21$ Результаты экспериментов представлены в таблице 3 и 4.

Время передачи данных между КС для КВС МЭИ Таблица 3

Объем передаваемых данных, Кб	1,6	8	48	80	800
время передачи данных (мкс) между вычислителями в одном ВУ	2,68	10,5	54,1	79,6	590
время передачи данных (мкс) между вычислителями в разных ВУ	10,1	26,3	88,1	133,6	1180

Время передачи данных между КС для КВС ХТУ Таблица 4

Объем передаваемых данных, Кб	1,2	12	60	120	240	1200
время передачи данных (мкс) между вычислителями в одном ВУ	2,5	15,9	77,49	88,7	166	780
время передачи данных (мкс) между вычислителями в разных ВУ	35,56	347,6	1707	1841,6	3825,2	16860

Результаты приведенные в таблицах показывают, что полученные коэффициенты для двух КВС имеют большие различия. Время на межузловые передачи данных для КВС МЭИ в 2 раза дольше, чем при передаче данных внутри узла, а для КВС ХТУ эта разница составляет 21 раз. Столь разные значения связано с техническими характеристиками КВС (КВС МЭИ использует технологию Infinband, а КВС ХТУ использует технологию Gigabit Ethernet). Определены коэффициенты накладных расходов $\varepsilon_2^{'}, \varepsilon_{2,}^{''} \varepsilon_2^{'''}$ при передаче данных на разных уровнях иерархии, которые возникают когда идет нагрузка на сетевой адаптер в ВУ. По результатам проведенных исследований автор предлагает сделать преобразование графа ПА с учетом влияния вышеперечисленных факторов на время выполнения вершин графа и время передачи данных. Данное представление графа алгоритма будет использоваться при решении задачи назначения. Пример результата преобразования графа представлен на рис. 2.

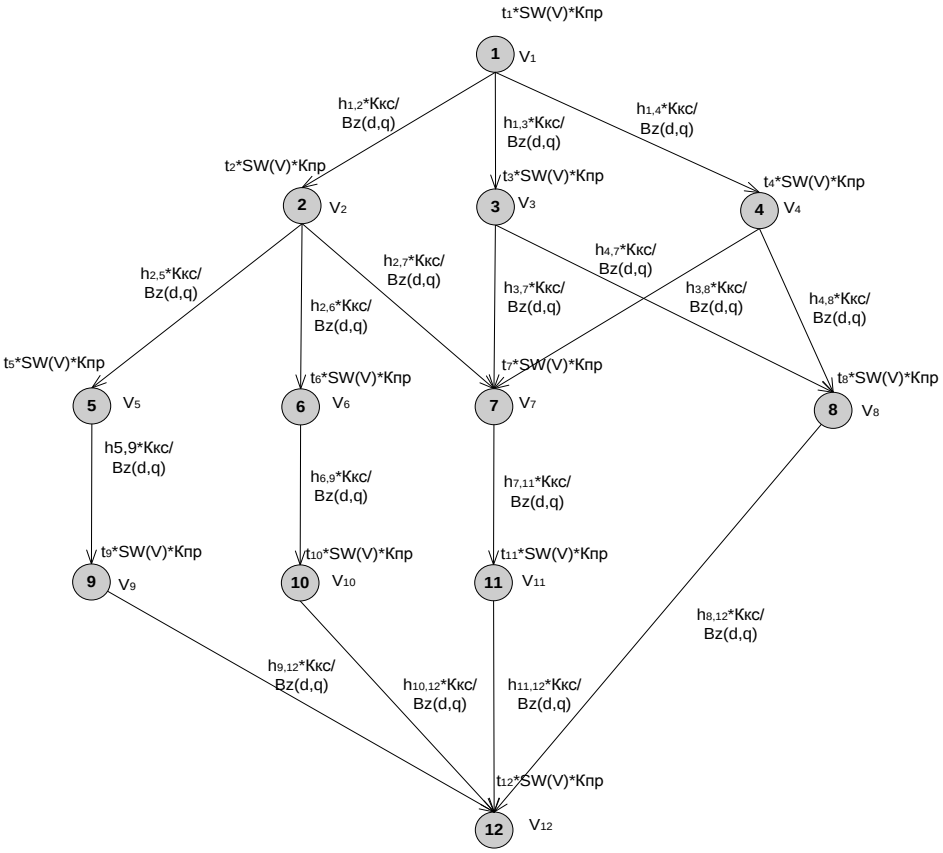

Рис. 2. Преобразованный граф алгоритма с учетом накладных расходов при использовании разных уровней иерархии КС и памяти

В кружках-вершинах графа стоят их номера. Вершины взвешены временем их выполнения t_i, на отдельно взятом вычислителе, и емкостью памяти V_i, требуемой для хранения данных соответствующего ФПП.

Кпр - коэффициент уменьшения производительности вычислителя, выполняющего *i*-ю вершину с участием модуля памяти *l*, принимает значение в зависимости от числа пользователей, использующих данный модуль памяти и значения коэффициента накладных расходов при обращении к данному

модулю памяти. Данный коэффициент отражает то, насколько увеличивается время выполнения вершины *i* при учете выше перечисленных значений.

$$Knp = U\ (PL) \times (1 + Eps(SW(V), Conf))$$

SW(V) - функция принимающая значение $\partial_1^{'}$ или $\partial_1^{''}$ или $\partial_1^{'''}$ в зависимости от объема обрабатываемых данных.

$U\ (PL)$ - коэффициент уменьшения производительности вычислителя, определяемый количеством пользователей, использующих модуль памяти l в ВУ.

$Eps(SW(V), Conf)$ коэффициент уменьшения производительности вычислителя, принимающий значение $\varepsilon_1^{'}$ или $\varepsilon_1^{''}$ или $\varepsilon_1^{'''}$ в зависимости от уровня иерархии доступа к памяти и количества вычислителей, использующих данный модуль память.

Третья глава посвящена разработке модели и алгоритма назначения фрагментов ПП на вычислители КВС, с целью минимизации времени выполнения ПП на заданных ресурсах КВС, за счет поиска эффективного варианта назначения. Для поиска такого варианта назначения ставится задача разработки оптимизационной модели и алгоритма эффективного назначения ФПП на вычислители КВС. Процесс решения оптимизационной задачи можно представить следующим образом на рис 3.

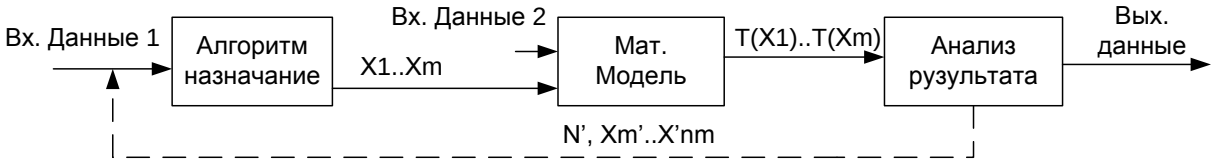

Рис. 3. Процесс решения оптимизационной задачи

Здесь Вх. Данные 1 – характеристики прикладной задачи (например, граф алгоритма представленный в виде матрицы связанности и т.д.), Вх. Данные 2 – характеристики модели КВС, *X*1..*Xm* – варианты матрицы назначений, *N′* – число циклов поиска оптимального результата.

Автором предлагается два варианта решения поставленной задачи.

1. Поярусная модель оптимизации. Благодаря этому способу можно выразить время выполнения ПП в виде аналитической зависимости от характеристик КВС и варианта назначения. При построении модели поярусной оптимизации делается следующее ограничение - переход к следующему ярусу происходит только тогда, когда выполнены все вершины данного яруса. Таким образом, при прохождении каждого яруса, происходит оценка верхней границы затрачиваемых ресурсов аппаратных средств. Такая оценка влияет на точность решения задачи назначения. Поэтому недостатком данного варианта является возможность его использования только для узкого класса задач.

2. Имитационное моделирование по интервалам времени с применением генетического алгоритма. Весь процесс моделирования разбивается на множество интервалов времени, и на каждом интервале моделируется состояние системы. Таким образом, можно зафиксировать моменты переходов

состояний системы и накладные расходы, которые необходимо учитывать при моделировании.

В разделе 3.2 рассматривается математическая модель общего вида (соответствующая первому варианту), записанная в аналитической форме и отражающая зависимость времени выполнения ПП от варианта назначения.

$$T(x)=\sum_{e=1}^{p}\left(\max_{1\le d\le K}\left\{\sum_{i=\gamma_{e\ \min}}^{\gamma_{e\ \max}}\left(\sum_{l=1}^{L}A_i\times xm_{i,l}\times x_{i,d}\times SW(V_{e,l})\times Knp_{l_i'}+\right.\right.\right.$$
$$\left.\left.\left.+\sum_{j=1}^{N}\sum_{d=1}^{K}\sum_{q=1}^{K}x_{i,d}\times x_{j,q}\times\frac{h_{i,j}}{B_{z(d,q)}}\right)\right\}\right)$$

$$V_{e,l}=\sum_{i=\gamma_{e\ \min}}^{\gamma_{e\ \max}}xm_{i,l}\times V_i^{out}+\sum_{d=1}^{K}mm_{d,l}\times\max_{\gamma_{e\ \min}\le i\le\gamma_{e\ \max}}\left\{x_{i,d}\times(V_i-V_i^{out}\right\}$$

$$\sum_{d=1}^{K}x_{i,d}=1,\ \sum_{q=1}^{K}x_{j,q}=1,\ i,j\in[1..N];\ d,q\in[1..K]$$

$$V_{e,l}-V_{free_l}-V_{swap_l}>0\rightarrow\min$$

Объяснение индексов используемых в математической модели:

$e\in\{1,..,P\}$, где P количество ярусов в графе.

$\gamma_{e\ \min}$ - номер младшей вершины на ярусе "e"

$\gamma_{e\ \max}$ - номер старшей вершины на ярусе "e"

матрица *H*, где h_{ij} объём передаваемых данных (байт) от вершины *i* к вершине *j*. *1<i,j<N*. из данной матрицы *H* можно найти объём данных передаваемых i-й вершиной всем последователям $V_i^{out}=\sum_{j=1}^{N}h_{i,j}$. Из матрицы *H* можно найти объём принимаемых данных для j-й вершины от предшественников $V_j^{in}=\sum_{i=1}^{N}h_{i,j}$.

L общее число модулей памяти в системе, где $l\in[1,..,L]$. V_{free} *[l]* – емкость свободной памяти в модуле *l*. V_{swap} *[l]* – ёмкость памяти на внешнем носителе, выделяемой КВС при свопинге для ВУ, к которому подключен модуль памяти *l*.

XM матрица назначения вершины на использование модуля памяти *l*, $xm_{i,l}=1$, если вершина i использует модуль памяти *l*, где $\forall i\in N;\forall l\in L$.

MM матрица, $mm_{d,l}$ = 1, если процессор *d* подключен к модулю *l*, 0 в противном случае, где $\forall d\in K;\forall l\in L$.

$VE_{e,l}$ суммарная критическая емкость памяти, которая может быть затребована у модуля *l (*для хранения выходных данных и дополнительных данных в процессе выполнения вершин) в процессе выполнения на ярусе “e”.

Данная аналитическая зависимость построена с учетом следующих ограничений. Считается, что в процессе решения задачи состояние КВС не меняется. В работе не рассматриваются вопросы, связанные с управлением процессами и не учитываются факторы, связанные с порождением процессов ОС и другими службами управления, которые тоже вносят задержки во время выполнения программ. КВС является гомогенной. Емкость оперативной и дисковой памяти достаточно для размещения данных программы. Передача данных осуществляется последовательно от младшей вершины к старшей. Передача данных осуществляется без конфликтов и простоев. Каждая вершина может назначаться только на один вычислитель.

Согласно построенной аналитической зависимости оптимизационная модель назначения имеет вид:

$$T(x)=\left\{\sum_{e=1}^{p}\left(\max_{1\le d\le K}\left\{\sum_{i=\gamma_{e\,min}}^{\gamma_{e\,max}}\left(\sum_{l=1}^{L}A_i\times xm_{i,l}\times x_{i,d}\times SW(V_{e,l})\times Knp_{l_i}+\right.\right.\right.\right.$$
$$\left.\left.\left.\left.+\sum_{j=1}^{N}\sum_{d=1}^{K}\sum_{q=1}^{K}x_{i,d}\times x_{j,q}\times\frac{h_{i,j}}{B_{z(d,q)}}\right)\right\}\right)\right\}\to\min$$

$$V_{e,l}=\sum_{i=\gamma_{e\,min}}^{\gamma_{e\,max}}xm_{i,l}\times V_i^{out}+\sum_{d=1}^{K}mm_{d,l}\times\max_{\gamma_{e\,min}\le i\le\gamma_{e\,max}}\left\{x_{i,d}\times(V_i-V_i^{out}\right\}$$

$$\sum_{d=1}^{K}x_{i,d}=1;\ \sum_{q=1}^{K}x_{j,q}=1;\ \ i,j\in[1..N];\ d,q\in[1..K]$$

$$V_{e,l}-V_{free_l}-V_{swap_l}>0\to\min$$

Критерием оптимизации является время выполнения ПП. В результате оптимизации должна быть получена матрица *X*, определяющая эффективное назначение ФПП на выделенные вычислители КВС. Решение задачи оптимизации на представленной модели назначения является весьма трудоёмким процессом, поэтому для «больших задач» применить модель невозможно. Обычно для реализации данной оптимизационной модели используют эвристические алгоритмы назначения и имитационное моделирование. В разделе 3.2.4 подробно описан принцип работы имитационной модели, а так же представлена схема алгоритма. В качестве алгоритма назначения был выбран генетический алгоритм (ГА). В разделе 3.3 подробно описан принцип работы и схема ГА. Одним из важных моментов для использования ГА является представление параметров объектов рассматриваемой предметной области исследований в виде генов особей. Для решения поставленной задачи в диссертации используются следующие представления.

1. Особь - носитель генетического кода, для нашей задачи это массив назначений вершины a_i графа ПА на вычислители КВС.
2. Генетический код - последовательность хромосом.
3. Хромосома - ячейка в массиве назначения вершин на вычислители КВС.

4. Потомок - особь, полученная путем комбинации значений родительских особей (массивов назначений).
5. Популяция - набор особей - для нашей задачи это множество массивов назначений.

Подробная схема ГА для задачи назначения представлен на рис. 4.

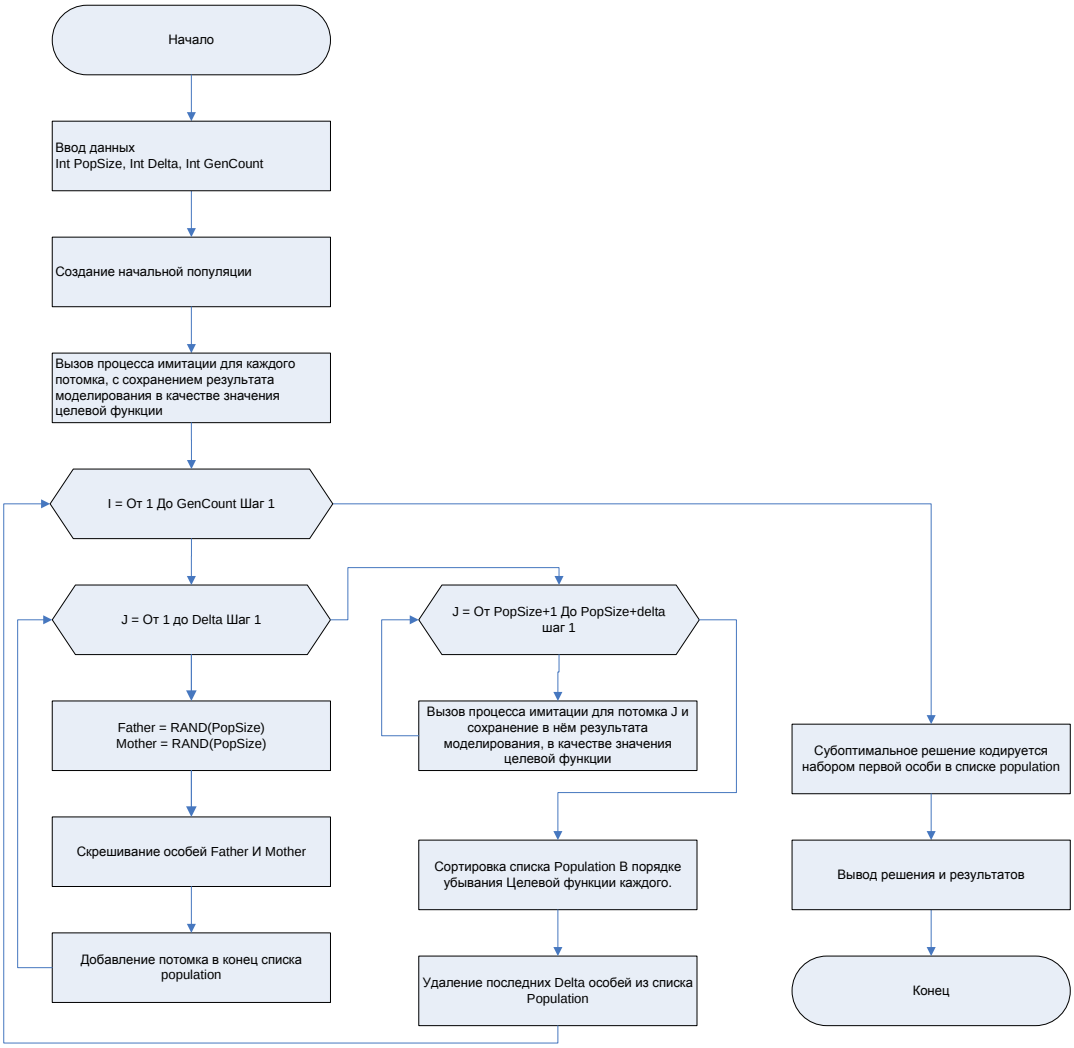

Рис. 4. Схема генетического алгоритма для задачи назначения

В работе было проведено сравнение полученных результатов решения задачи назначения, показывающих зависимость времени выполнения алгоритма от найденного варианта эффективного назначения $T(X)$ для разных модельных задач с помощью ГА назначения и при полном переборе. Результат сравнения показал, что найденное время решения задачи при применении ГА приближенно равно минимальному времени решения задачи при полном переборе для всех вариантов назначения, а время, затрачиваемое на поиск эффективного варианта назначения существенно меньше, чем при полном переборе. Выигрыш во времени достигается из-за того, что для поиска минимального времени решения задачи количество переборов при ГА на порядок меньше, чем при полном переборе. Результат представлен в табл. 5

Время выполнения задачи при использовании ГА и полного перебора Таблица 5

Число используемых вычислителей	3	3	3	3	3	3	3	3	4	4
Расположение используемых вычислителей в КВС	все 3 в одном ВУ	2 в одном ВУ,1 в другой ВУ	все 3 в разные ВУ	все 3 в одном ВУ	2 в одном ВУ,1 в другой ВУ	все 3 в разных ВУ	2 в одном ВУ,1 в другой ВУ	все 3 в одном ВУ	3 в одном ВУ,1 в другой ВУ	3 в одном ВУ,1 в другой ВУ
Номер графа алгоритма	1	1	1	2	2	2	3	3	4	5
Количество вершин /дуг	12/15	12/15	12/15	13/19	13/19	13/19	14/21	14/21	14/19	13/20
Кол-во матриц назначения для решения задачи при полном переборе K^N	531441	531441	531441	1594323	1594323	1594323	4782969	4782969	268435436	67108864
Кол-во вариантов матриц назначений при ГА	11000	11000	11000	11000	11000	11000	6000	6000	51000	51000
Минимальное теор. Время решения(Ткр)	23	23	23	160	160	160	190	190	218	212
Максимальное теор. время решения(Тпосл)	48	48	48	370	370	370	460	460	440	400
Минимальное время решения задачи при полном переборе	39	37	37	209	193	197	226	247	*	*
Минимальное время решения задачи при ГА	37	37	39	210	217	211	246	251	231	228

В работе исследовались различные варианты сочетания параметров ГА для проверки достоверности полученных результатов. Так как ГА является эвристическим алгоритмом, то для достоверности полученных результатов был рассчитан доверительный интервал на модельных задачах, Для задачи № 2 из таблицы 5, где все 3 вычислителя расположены в одном ВУ при доверительной вероятности равной 0,95 и при выборке равной тридцати, доверительный интервал составил (208, 212). Для задачи № 3 из таблицы 5, где все 3 вычислителя расположены в одном ВУ, доверительный интервал составил (250, 254). Полученный результат показал узкий диапазон сходимости результатов, что подтверждает правильность работы ГА. Для доказательства адекватности разработанной модели было проведено сравнение реального времени решения ПП на КВС и времени решения данной ПП на модели для разных вариантов матриц назначениях $X1$ и $X2$.

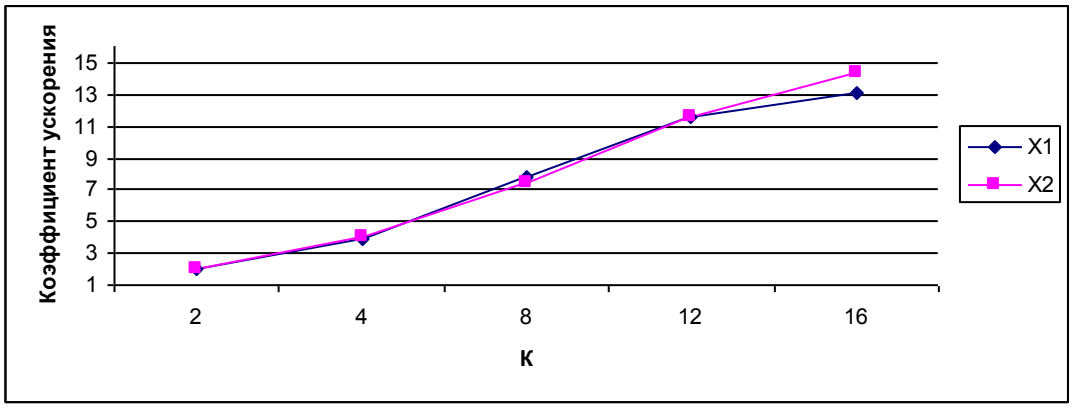

Рис.5 Зависимость коэффициента ускорения от количества вычислителей и матрицы назначения на КВС.

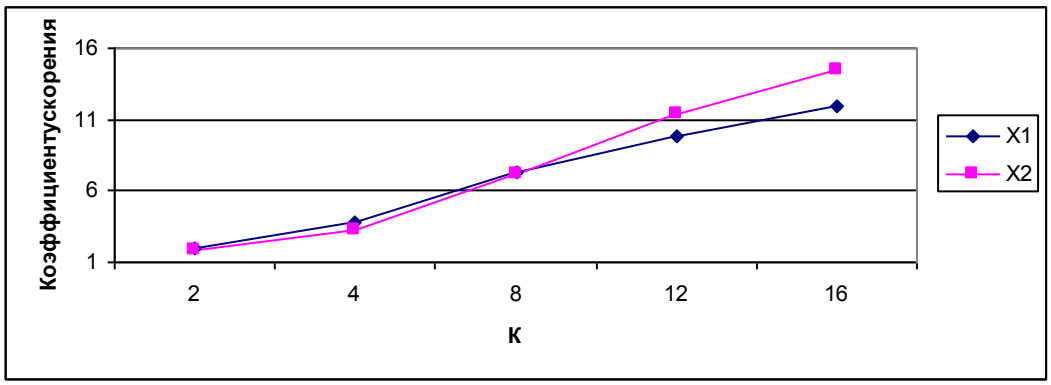

Рис.6 Зависимость коэффициента ускорения от количества вычислителей и матрицы назначения на модели.

Время решения ПП на модели показало одинаковый динамический характер изменения времени решения ПП от разных вариантов назначений. Для иллюстрации этого факта представлены зависимости коэффициентов ускорения от числа используемых вычислителей при решении задачи перемножения матриц при разных вариантах назначения на КВС на рис. 5 и на модели рис 6.

Применение разработанного алгоритма назначения фрагментов ПП позволяет получить вариант эффективного назначения быстрее, чем при полном переборе.

Таким образом, основные результаты, полученные в третьей главе, состоят в разработке оптимизационной модели и алгоритма назначения ФПП на выделенные вычислители КВС, учитывающего особенности иерархической организации памяти, КС и накладные расходы для минимизации времени выполнения ПП на кластерах.

В **четвертой главе** приведены критерии для анализа эффективности разработанного алгоритма назначения по сравнению с другими алгоритмами или полным перебором. Это коэффициент ускорения, коэффициент загрузки ресурсов КВС и коэффициент, отражающий отношение времени выполнения ПП, полученное с помощью разработанного алгоритма назначения, ко времени полученному другими алгоритмами. В разделе 4.2 приводятся примеры анализа эффективности на модельных задачах. На рис 7. показаны зависимости коэффициента ускорения и коэффициента загрузки для модельной задачи от разных вариантов матриц назначения: (ГА), найденная с помощью ГА назначения, (СЛ), алгоритма случайного выбора и (Перебор) полным перебором. Модельная задача относится к задачам, в которых суммарное время передачи данных больше суммарного времени выполнения вершин. Как показано на рис.7, как только количество вычислителей превышает 4-х, коэффициент ускорения начинает падать из-за временных затрат на передачи данных между вычислителями.

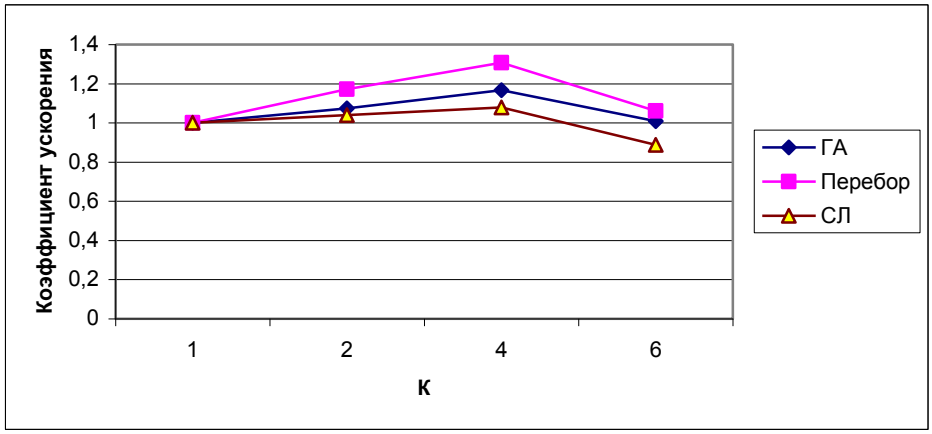

Рис.7 Зависимость коэффициента ускорения от количества вычислителей и алгоритма назначения.

Также для анализа эффективности разработанного алгоритма проведены эксперименты, где в качестве прикладных задачах использовались следующие.

1. ПП задачи перемножения матриц,
2. ПП анализа надежности[1] методом моделирования по интервалам времени.
3. Параллельная MPI программа расчета освещенности виртуальных экранов (3D–рефрактограмм) и визуализации освещенности экрана в интерференционной схеме для модели сферически неоднородной среды. Данная прикладная задача широко используется в области оптических методов измерения, цель которых заключается в исследование процесса остывания нагретого объекта бесконтактным методом.

Экспериментальные исследования показали, что конфликты на одновременное обращение к памяти влияют на эффективность выполнения параллельных программ, относящихся к слабосвязанным классам задач (т.е. время выполнения ФПП намного больше, чем время передачи данных между фрагментами), такие как задача перемножения матриц или задача теплопроводности[1]. На рис. 8 представлена зависимость коэффициента ускорения от количества вычислителей и матриц назначения ГА и Х при решении задачи перемножения матриц на КВС МЭИ (матрица Х получена с помощью эвристического метода, основанного на минимизации времени передачи данных). На рис. 9 представлена аналогичная зависимость, что и на рис. 8, при выполнении задачи на КВС МГУ и ХТУ (матрица Х получена по принципу, заложенному в планировщике КВС)

[1] Автореферат. Яньков Сергей Георгиевич. Исследование и разработка методики отображения задач на кластерные системы с иерархически-неоднородной коммуникационной средой .

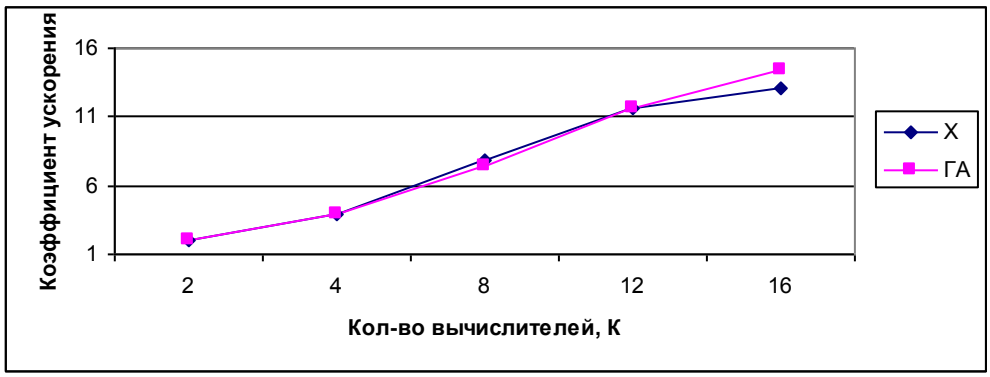

Рис. 8. Зависимость коэффициента ускорения от количества вычислителей и алгоритма назначения.

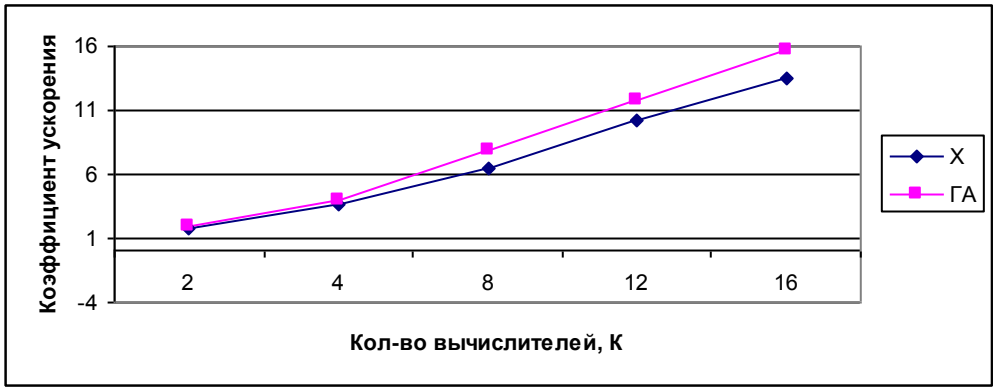

Рис. 9. Зависимость коэффициента ускорения от количества вычислителей и алгоритма назначения.

При решении задачи моделирования процесса остывания нагретого объекта на КВС МЭИ и МГУ, несмотря на различные варианты назначения, коэффициент ускорения имеет одинаковое значение рис.10. Данный результат объясняется характеристикой данной задачи, объем обрабатываемых данных в задаче минимален и помещается в КЭШ памяти, следовательно, конфликты и накладные расходы на обращения к памяти почти отсутствуют. Интенсивность передачи данных между ФПП минимальна и не соизмерима с временем выполнения ФПП, следовательно, накладные расходы при передачи данных или иерархия КС не могут повлиять на время выполнения ПП.

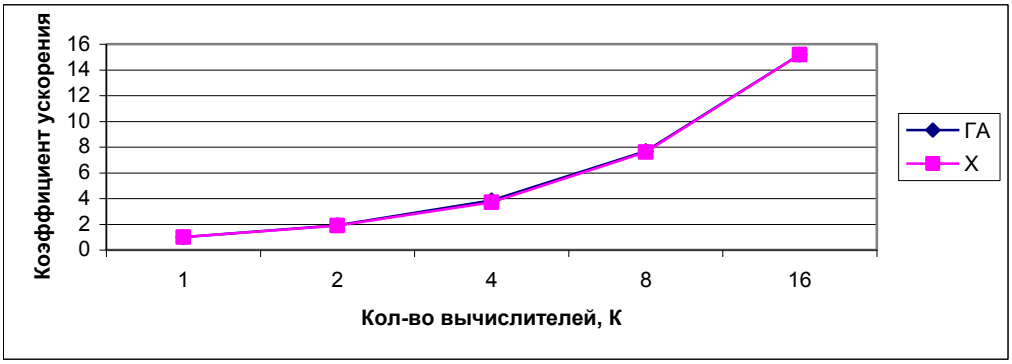

Рис.10 Зависимость коэффициента ускорения от количества вычислителей и алгоритма назначения на КВС МЭИ

Для подтверждения влияния конфликтов при передаче данных между вычислителями, был проведен анализ выполнения задачи «Надежность». Полученный результат показал, что конфликты при передаче данных влияют на эффективность выполнения ПП. На рис.11 представлена зависимость коэффициента ускорения от количества вычислителей и матриц назначения, полученная по ГА или другими эвристическими алгоритмами X1 и X2.

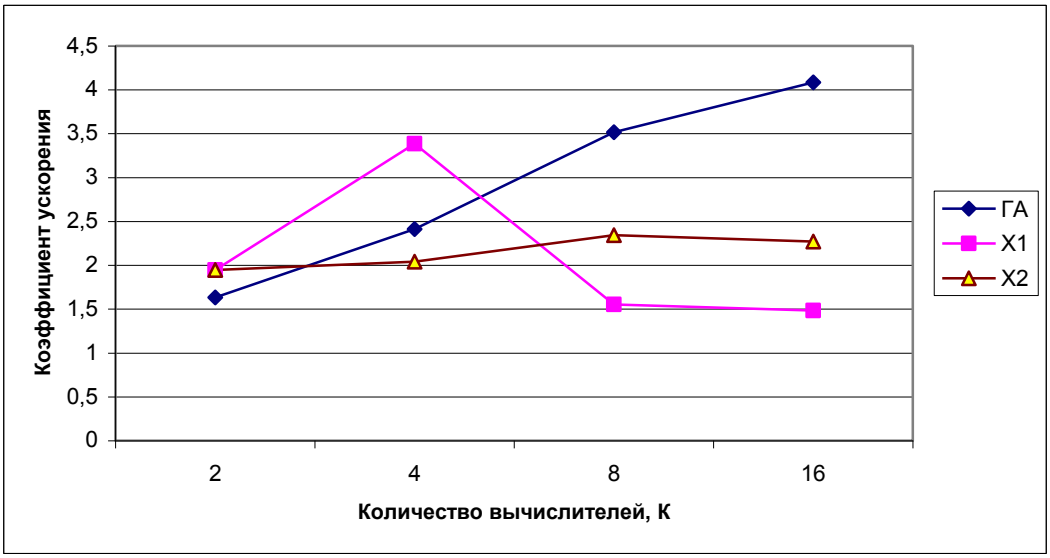

Рис.11 Зависимость коэффициента ускорения от количества вычислителей и алгоритма назначения

Таким образом, применение разработанного алгоритма назначения позволило повысить эффективность выполнения ПП на КВС рассматриваемых задач на 6-12% по сравнению с принципом назначения, заложенным в планировщике КВС или в существующих эвристических методах назначения.

В диссертации показано, что учет иерархической неоднородной организации КС, памяти и накладных расходов для задачи назначения может повлиять на время выполнения ПП. Для эффективного выполнения ПП необходимо провести предварительное тестирование для определения значений введенных коэффициентов. На основе проведенных исследований и разработанного алгоритма назначения, автором предложена методика назначения ФПП на вычислители КВС. Методика состоит из следующих этапов.

1. Предварительная оценка эффективности выполнения параллельных программ.
2. Формальное представление решаемой задачи и КВС.
3. Тестирование КВС для получения характеристик, необходимых для имитационной модели.
4. Назначение фрагментов ПП на вычислители КВС с помощью разработанного алгоритма.
5. .Оценку результатов выполнения ПП на КВС.

Процедуры выполнения всех вышеперечисленных этапов определяют содержание разработанной методики назначения ПП на КВС.

В **заключении** кратко сформулированы основные результаты, полученные в диссертационной работе.

1. Проведен анализ современных КВС, построенных на разных платформах и выявлены особенности, влияющие на время выполнения параллельных программ.
2. Проведен обзор использующихся в настоящее время методов назначения ФПП и анализ их особенностей.
3. Разработана оптимизационная модель назначения на основе ГА для рассматриваемого в работе класса задач на вычислители КВС, учитывающая особенности их иерархически-неоднородной организации и накладные расходы на разных уровнях иерархии, для минимизации времени выполнения параллельных программ на современных КВС.
4. На основе разработанного алгоритма назначения, создана инструментальная программная система. Использование данной программной системы позволяет улучшить процесс выбора эффективного варианта назначения для заданных классов задач.
5. На основания проведенных в диссертации исследований и разработанного алгоритма назначения, была разработана методика назначения ФПП на вычислители КВС.
6. Проведены экспериментальные исследования эффективности разработанного ГА назначения на КВС МЭИ, МГУ, ХТУ. Полученный результат продемонстрировал повышение эффективности выполнения ПП на 6-12% по сравнению с алгоритмом назначения, заложенным в планировщике КВС, или другими эвристическими методами.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Во Минь Тунг. Исследование методов распределения данных по процессорам кластерной системы для заданного класса прикладных задач. // Труды международной научно-технической конференции «Информационные средства и технологии». 2008 . — С. 118-123.

Во Минь Тунг. Генетические алгоритмы в задаче о назначении. // Шестнадцатая международная научно-техническая конференция студентов и аспирантов . 2010 г. — С. 401 -402.

Во Минь Тунг. Оценка характеристик кластерных вычислительных систем. // Вестник МЭИ. — М.: Издательский дом МЭИ, 2010. — №2. — С. 133-140.

Подписано к печати Зак. Тир. 100 Печ.л.
Отпечатано в Полиграфическом центре МЭИ (ТУ) Красноказарменная ул., д. 13