На правах рукописи

Павлов Павел Александрович

ОДИНАКОВО РАСПРЕДЕЛЕННЫЕ СИСТЕМЫ КОНКУРИРУЮЩИХ ПРОЦЕССОВ

Специальность 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени кандидата физико-математических наук

Минск – 2008

Работа выполнена в Учреждении образования «Белорусский государственный экономический университет».

Научный руководитель: доктор физико-математических наук, профессор
Коваленко Николай Семенович

Официальные оппоненты: доктор физико-математических наук
Кузюрин Николай Николаевич

кандидат физико-математических наук
Бахмуров Анатолий Геннадьевич

Ведущая организация: *Московский энергетический институт*

Защита состоится 15 февраля 2008 года в 15 часов на заседании диссертационного совета Д 002.087.01 в Институте системного программирования Российской академии наук по адресу: 109004, Москва, ул. Большая Коммунистическая, д.25.

С диссертацией можно ознакомиться в библиотеке Института системного программирования РАН.

Автореферат разослан 13 января 2008 года.

Ученый секретарь
диссертационного совета
канд. физ.-мат. наук С.П. Прохоров

Общая характеристика работы

Актуальность темы

Необходимость достижения сверхвысокой производительности и надежности вычислительных средств, существенного ускорения решения реальных задач большой размерности и повышения точности результатов, неразрывно связана с созданием многопроцессорных систем сложной архитектуры. Создание таких систем и соответствующего программного обеспечения требует решения трудных в математическом отношении задач организации большого числа одновременно взаимодействующих параллельных процессов, расчета характеристик многопроцессорных вычислительных систем, распараллеливания алгоритмов, разработки приемов ускорения вычислений.

Среди перечисленных задач следует выделить задачи распределения ограниченных вычислительных ресурсов в условиях конкуренции за их использование. В связи с этим актуальным является дальнейшее развитие математической модели распределенной обработки одновременно взаимодействующих конкурирующих процессов, решение задач организации выполнения распределенных процессов в базовых режимах, сравнительного анализа режимов, поиска условий эффективности и оптимальной организации выполнения распределенных процессов. Один из подходов на пути решения указанных задач основывается на структурировании программных ресурсов на параллельно выполняемые блоки с их последующей конвейеризацией по процессам и процессорам.

В работах Иванникова В.П., Коваленко Н.С., Метельского В.М. введена математическая модель распределенной обработки конкурирующих процессов, определены базовые режимы взаимодействия конкурирующих процессов, процессоров и блоков, введены определения неоднородной, однородной и одинаково распределенной систем конкурирующих процессов, получены математические соотношения для вычисления точных значений минимального общего времени выполнения заданных объемов вычислений в случаях неограниченного параллелизма по числу процессоров многопроцессорной системы, получены критерии эффективности и оптимальности структурирования программных ресурсов. В то же время остаются нерешенными следующие задачи:

- нахождения минимального общего времени выполнения множества распределенных конкурирующих процессов с учетом дополнительных системных расходов;
- сравнительного анализа асинхронного и двух синхронных режимов организации процессов при распределенной обработке;
- определения необходимых и достаточных условий эффективности и оптимальности одинаково распределенных систем конкурирующих процессов в различных режимах их взаимодействия в условиях неограниченного и ограниченного параллелизма.

Цель и задачи работы

Целью диссертационной работы является дальнейшее развитие математической модели организации распределенных вычислений над структурированными программными ресурсами и решение на ее основе дискретно–комбинаторных оптимизационных задач, возникающих при выполнении одинаково распределенных конкурирующих процессов, проведение сравнительного анализа времен реализации распределенных конкурирующих процессов для асинхронного и двух синхронных режимов.

В соответствии с поставленной целью в диссертации детально исследованы следующие задачи:

- определения минимального общего времени реализации распределенных конкурирующих процессов в различных режимах взаимодействия процессов, процессоров и блоков программного ресурса с учетом дополнительных системных расходов;
- сравнительного анализа временных характеристик реализации одинаково распределенных конкурирующих процессов в базовых режимах;
- определения необходимых и достаточных условий эффективности и оптимальности одинаково распределенных систем конкурирующих процессов.

Научная новизна

Научной новизной обладают следующие результаты диссертационной работы:

- решены задачи нахождения минимального общего времени выполнения неоднородных, однородных и одинаково распределенных конкурирующих процессов в асинхронном и синхронных режимах с учетом дополнительных системных расходов [1–2, 5, 8–10, 13–14];

▪ проведен сравнительный анализ времен реализации в базовых режимах одинаково распределенных конкурирующих процессов при достаточном числе процессоров многопроцессорной системы [3, 5–6];

▪ в случае неограниченного параллелизма для всех трех базовых режимов установлены достаточное условие эффективности одинаково распределенных систем конкурирующих процессов и необходимое и достаточное условие существования эффективных систем одинаково распределенных конкурирующих процессов в зависимости от величины дополнительных системных расходов [1, 3, 10, 15];

▪ в случае ограниченного параллелизма условия эффективности одинаково распределенных систем конкурирующих процессов получены для асинхронного и второго синхронного режимов [1, 3, 10, 15];

▪ в условиях неограниченного и ограниченного параллелизма решены задачи оптимальности одинаково распределенных систем конкурирующих процессов [4, 7, 11].

Практическая значимость

Проведенные исследования позволяют давать практические рекомендации по организации распределенных процессов, конкурирующих за использование общих программных ресурсов в различных режимах их взаимодействия применительно к вычислительным системам с распределенной обработкой данных. В частности, предложенные методы и формулы позволяют построить расписания моментов запуска и окончания каждого из конкурирующих процессов, что дает возможность решать проблему синхронизации процессов, существенно минимизировать накладных расходы, свести к минимуму непроизводительные простои процессоров и задержки выполнения блоков. Характер полученных формул позволяет также явно учитывать накладные расходы времени, связанные с затратами на реализацию механизмов управления параллельными процессами при распределенной обработке. Решенные задачи позволяют исследовать всевозможные смешанные режимы организации выполнения параллельных процессов при распределенной обработке, в том числе с учетом ограниченного числа копий структурированного программного ресурса. Полученные математические соотношения служат основой для решения оптимизационных задачи распределенных вычислений.

Апробация результатов диссертации

Диссертационная работа выполнялась в рамках Белорусского республиканского фонда фундаментальных исследований по теме «Разработка новых подходов к распараллеливанию численных методов математической физики на основе анализа тонких свойств графов», проект № Ф04Р–156, время выполнения с 01.05.2004 г. по 01.06.2006 г.

Все основные результаты, включенные в диссертационную работу, получены автором лично. Соавторы совместных публикаций приняли участие в постановке задач, обсуждении полученных результатов.

Материалы, вошедшие в диссертационную работу, докладывались и обсуждались на научных семинарах в отделе рекурсивных вычислений Института кибернетики АН Украины и Институте математики НАН Беларуси, а также на 10 научно–практических конференциях: республиканской научной конференции «Социально–экономическое и гуманитарное развитие белорусского общества в XXI веке», секция «Информационные технологии и математические методы в экономике» (Минск, 2004); научно–практической конференции «Экономический механизм формирования национальной модели развития экономики РБ», секция «Развитие информационных технологий в Республике Беларусь» (Пинск, 2005); II научно–практической конференции «Актуальные проблемы рыночной экономики», секция «Статистические и математические методы в экономике» (Бобруйск, 2005); международной научной конференции «Экономическое развитие Беларуси в контексте расширения Европейского Союза», секция «Информационные технологии в экономических процессах» (Гродно, 2005); II научно–практической конференции «Исследования молодых ученых Пинщины», секция «Информационных технологий и компьютерные коммуникации» (Пинск, 2005); научно–практической конференции «Механизм формирования социально–экономического развития регионов Республики Беларусь в условиях перехода к рыночной экономике», секция «Внедрение информационных технологий в повышение эффективности социально–экономического развития Белорусского Полесья» (Пинск, 2006); международной научно–практической конференции «Механизмы устойчивого развития инновационных социально–экономических систем», секция «Статистические и математические методы в экономике» (Бобруйск, 2006); международной научно-практической конференции «Социально-экономическое и историко-культурное развитие Полесского региона в XXI веке», секция «Техника, информационных технологий и компьютерные коммуникации» (Пинск, 2006); VII международной конференции «Проблемы прогнозирования и государственного регулирования социально–экономического развития», секция «Математическое регулирование экономических процессов» (Минск, 2006); третьей международной конференции «Информационные системы и технологии

IST'2006», секция «Параллельная и распределенная обработка данных» (Минск, 2006).

Опубликованность результатов

Основные результаты диссертации опубликованы в 16 работах. Статей в научных и научно–теоретических журналах 6. Количество опубликованных в них материалов 1,42 авторских листа. Тезисов докладов и выступлений на международных и республиканских научных конференциях 9.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения и библиографического списка из 85 наименований. Объем диссертационной работы составляет 91 страницу машинописного текста, включая 13 рисунков.

Краткое содержание работы

Во **введении** дается краткая оценка современного состояния проблемы, рассматриваемой в диссертации, очерчивается круг задач, нуждающихся в изучении, определяется направление диссертационного исследования.

В **первой** главе в 1.1 проведен обзор литературы, кратко освещены работы предшественников по задачам, рассматриваемым в работе.

В 1.2 конкретизированы понятия *процесса* и *программного ресурса*, поскольку именно они являются конструктивными единицами моделей функционирования программ, реализующих методы распределенной обработки решения задач. *Процесс* в работе рассматривается как последовательность команд (блоков) $I_s = (1, 2, ..., s)$. При этом процесс называется *распределённым*, если все блоки или часть из них выполняются на разных процессорах. *Программный ресурс* определяется как многократно выполняемая в многопроцессорной системе программа или ее часть. Используя введенные понятия, в диссертации показано, что создание и исследование программ, использующих принципы распределенной обработки, сводится к решению задач организации взаимодействия распределенных процессов, конкурирующих за программный ресурс.

Методу *структурирования* программных ресурсов, который приобретает особенно фундаментальный характер в области распределенного программирования, посвящен 1.3. Основная идея данного метода, состоит в обеспечении структурирования программного ресурса на блоки Q_1, Q_2, ..., Q_s, с последующей конвейеризацией как блоков по процессам, так и процессов по процессорам многопроцессорной вычислительной системы, что позволяет получать существенный выигрыш по времени реализации заданных объемов вычислений.

В 1.4 вводится математическая модель распределенной обработки конкурирующих процессов, которая включает в себя p, $p \geq 2$, процессоров многопроцессорной системы, n, $n \geq 2$, конкурирующих процессов, s, $s \geq 2$, блоков структурированного на блоки программного ресурса, матрицу $T = [t_{ij}]$, $i = \overline{1,n}$, $j = \overline{1,s}$, времен выполнения блоков программного ресурса конкурирующими процессами. Предполагается, что все процессы используют одну копию структурированного на блоки программного ресурса, причем из физических соображений на множестве блоков установлен линейный порядок их выполнения. В исследовании вводится в рассмотрение параметр $\varepsilon > 0$, характеризующий время дополнительных системных (накладных) расходов, связанных с организацией параллельного использования блоков структурированного программного ресурса множеством конкурирующих процессов при распределенной обработке.

Предполагается, что взаимодействие процессов, процессоров и блоков программного ресурса подчинено следующим условиям:

1) ни один из блоков программного ресурса не может обрабатываться одновременно более чем одним процессором;

2) ни один из процессоров не может обрабатывать одновременно более одного блока;

3) обработка каждого блока программного ресурса осуществляется без прерываний;

4) распределение блоков программного ресурса по процессорам для каждого из процессов осуществляется циклически по правилу: блок с номером $j = kp + i$, $j = \overline{1,s}$, $i = \overline{1,p}$, $k \geq 0$ распределяется на процессор с номером i.

Введением дополнительных условий, определяются режимы взаимодействия процессов, процессоров и блоков. Условия 1–4 и условие

5) отсутствуют простои процессоров при условии готовности блоков, а также отсутствует невыполнение блоков при наличии процессоров, определяют *асинхронный* режим.

Условия 1–4 и условие

6) для каждого из n процессов момент завершения выполнения j-го блока на i-м процессоре совпадает с моментом начала выполнения

следующего $(j+1)$-го блока на $(i+1)$-м процессоре, $i=\overline{1,p-1}$, $j=\overline{1,s-1}$, определяют *первый синхронный* режим, который обеспечивает непрерывное выполнение блоков программного ресурса внутри каждого из процессов, а условия 1–4 и условие

7) для каждого из блоков момент завершения его выполнения l-м процессом совпадает с моментом начала его выполнения $(l+1)$-м процессом на том же процессоре, $l=\overline{1,n-1}$, определяют *второй синхронный* режим, обеспечивающий непрерывное выполнение каждого блока всеми процессами.

Определение 1.1. Система n распределенных конкурирующих процессов называется *неоднородной*, если времена выполнения блоков программного ресурса Q_1, Q_2, ..., Q_s зависят от объемов обрабатываемых данных и/или их структуры, т. е. разные для разных процессов.

Определение 1.2. Систему распределенных конкурирующих процессов будем называть *однородной*, если времена выполнения Q_j-го блока каждым из i–х процессов равны, т. е. $t_{ij}=t_j$, $i=\overline{1,n}$, $j=\overline{1,s}$.

Определение 1.3. Систему конкурирующих процессов будем называть *одинаково распределенной*, если времена t_{ij} выполнения блоков Q_j, $j=\overline{1,s}$, программного ресурса каждым из i-х процессов совпадают и равны t_i для всех $i=\overline{1,n}$, т. е. справедлива цепочка равенств $t_{i1}=t_{i2}=...=t_{is}=t_i$ для всех $i=\overline{1,n}$.

Вторая глава посвящена решению задач определения минимального общего времени выполнения множества распределенных конкурирующих процессов в каждом из режимов. Здесь же решена задача сравнительного анализа режимов с точки зрения временных затрат с учетом накладных расходов ε.

В 2.1 рассматривается *асинхронный* режим. Минимальное общее время выполнения *неоднородных* распределенных конкурирующих процессов $T_H^{ac}(p,n,s,\varepsilon)$, т. е. когда времена выполнения блоков программного ресурса разные для разных процессов, в случае *достаточного* числа процессоров $2\le s\le p$, и с учетом параметра $\varepsilon>0$, определяется по формуле:

$$T_н^{ac}(s,n,s,\varepsilon)=(n+s-1)\varepsilon+\max_{1\le u_1\le u_2\le...\le u_{s-1}\le n}\left[\sum_{i=1}^{u_1}t_{i1}+\sum_{i=u_1}^{u_2}t_{i2}+...+\sum_{i=u_{s-1}}^{n}t_{is}\right],$$

где u_1, u_2, ..., u_{s-1} – целые числа.

Задача нахождения $T_н^{ac}(p,n,s,\varepsilon)$ в 2.1 решена и с помощью аппарата *сетевых вершинно-взвешенных* графов. В частности, минимальное общее время в случае *неограниченного параллелизма* ($2 \le s \le p$), определяется длиной критического пути из начальной вершины в конечную. Каждой вершине графа соответствует значение $t_{ij}^{\varepsilon} = t_{ij} + \varepsilon$, $i = \overline{1,n}$, $j = \overline{1,s}$. Дуги в графе отражают линейный порядок выполнения блоков Q_j, $j = \overline{1,s}$, программного ресурса каждым из процессов, а также линейный порядок использования одних и тех же блоков разными процессами.

В случае *ограниченного параллелизма*, т. е. когда $s > p$, $s = kp + r$, $k \ge 1$, $1 \le r < p$, исходная матрица времен выполнения блоков с учетом дополнительных системных расходов $T^{\varepsilon} = [t_{ij}^{\varepsilon}]$, $i = \overline{1,n}$, $j = \overline{1,kp+r}$, разбивается на *(k+1)*–у подматрицу T_l^{ε}, $l = \overline{1,k+1}$, размерностью $n \times p$ каждая. По подматрицам T_l^{ε}, $l = \overline{1,k+1}$, строится результирующая матрица T^* размерностью $(k+1)n \times (k+1)p$, которая является блочной, симметричной, верхнедиагональной относительно второй диагонали, типа Ганкелевой порядка *k+1*.

$$T^* = \begin{bmatrix} T_1^{\varepsilon} & T_2^{\varepsilon} & T_3^{\varepsilon} & \cdots & T_k^{\varepsilon} & T_{k+1}^{\varepsilon} \\ T_2^{\varepsilon} & T_3^{\varepsilon} & T_4^{\varepsilon} & \cdots & T_{k+1}^{\varepsilon} & 0 \\ T_3^{\varepsilon} & T_4^{\varepsilon} & T_5^{\varepsilon} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ T_k^{\varepsilon} & T_{k+1}^{\varepsilon} & 0 & \cdots & 0 & 0 \\ T_{k+1}^{\varepsilon} & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Доказано, что минимальное общее время $T_H^{ac}(p,n,s,\varepsilon)$ выполнения *неоднородных* распределенных конкурирующих процессов в асинхронном режиме в случае $s = kp + r$, $k \ge 1$, $1 \le r < p$, определяется длиной критического пути из начальной вершины t_{11}^{ε} в конечную вершину $t_{(k+1)n,(k+1)p}^{\varepsilon}$ сетевого вершинно-взвешенного графа, веса которого задаются матрицей T^*.

Для *однородной* системы распределенных конкурирующих процессов, когда времена выполнения Q_j-го блока каждым из *i*-х процессов равны, т. е. $t_{ij}^{\varepsilon} = t_j^{\varepsilon}$, $i = \overline{1,n}$, $j = \overline{1,s}$, минимальное общее время в случае

$2 \le s \le p$, в асинхронном режиме составляет величину $T_o^{ac}(p,n,s,\varepsilon)$ равную

$$T_o^{ac}(p,n,s,\varepsilon) = T_\varepsilon^s + (n-1)\max_{1\le j\le s} t_j^\varepsilon ,$$

где $(t_1^\varepsilon,\ t_2^\varepsilon,\ \dots,\ t_s^\varepsilon)$ – длительности выполнения каждого из блоков программного ресурса с учетом накладных расходов, а $T_\varepsilon^s = \sum_{j=1}^{s} t_j^\varepsilon$ – длительность выполнения всего программного ресурса.

Для системы *одинаково распределенных* конкурирующих процессов, когда справедлива цепочка равенств $t_{i1}^\varepsilon = t_{i2}^\varepsilon = \dots = t_{is}^\varepsilon = t_i^\varepsilon$, $t_i^\varepsilon = t_i + \varepsilon$, для всех $i = \overline{1,n}$, справедлива

Теорема 2.1. *Минимальное общее время выполнения n, $n \ge 2$, одинаково распределенных конкурирующих процессов, использующих структурированный на s, $s \ge 2$, блоков программный ресурс в многопроцессорной системе с p, $p \ge 2$, процессорами и дополнительными системными расходами $\varepsilon > 0$, в асинхронном режиме составляет величину $T_{op}^{ac}(p,n,s,\varepsilon)$, равную*

$$T_{op}^{ac}(p,n,s,\varepsilon) = \begin{cases} T_\varepsilon^n + (s-1)\max\limits_{1\le i\le n} t_i^\varepsilon, & s \le p, \\ kT_\varepsilon^n + (p-1)\max\limits_{1\le i\le n} t_i^\varepsilon, & s = kp,\ k > 1, \\ (k+1)T_\varepsilon^n + (r-1)\max\limits_{1\le i\le n} t_i^\varepsilon, & s = kp + r,\ k \ge 1,\ 1 \le r < p, \end{cases}$$

где $T_\varepsilon^n = \sum_{i=1}^{n} t_i^\varepsilon$ – суммарное время выполнения каждого из блоков Q_j, $j = \overline{1,s}$, всеми n процессами.

В 2.2 решаются задачи нахождения минимального общего времени в *первом синхронном* режиме, при котором обеспечивается непрерывное выполнение блоков программного ресурса каждым из процессов. Как и для асинхронного режима, исследование проводится для неоднородных, однородных и одинаково распределенных процессов.

В случае, когда число процессоров многопроцессорной системы является *достаточным*, т. е. $s \le p$, для вычисления величины минимального общего времени выполнения *неоднородных* процессов в первом синхронном режиме $T_н^1(p,n,s,\varepsilon)$, имеет место формула:

$$T_н^1(p,n,s,\varepsilon) = \sum_{i=1}^{n-1} \max_{1 \le u \le s} \left[\sum_{j=1}^{u} t_{ij}^{\varepsilon} - \sum_{j=1}^{u-1} t_{i+1,j}^{\varepsilon} \right] + \sum_{j=1}^{s} t_{nj}^{\varepsilon}.$$

В случае, когда $s = kp$, $k > 1$, выполняется разбиение матрицы времен выполнения блоков T^{ε} на k подматриц по p столбцов в каждой. По каждой из подматриц строится линейная диаграмма Ганта. Воспользовавшись приемом совмещения последовательных диаграмм по оси времени, для вычисления минимального общего времени $T_н^1(p,n,kp,\varepsilon)$ в первом синхронном режиме получено выражение:

$$T_н^1(p,n,kp,\varepsilon) = \sum_{l=1}^{k} T_l^{\varepsilon} - \sum_{l=1}^{k-1} \delta_l ,$$

где $T_l^{\varepsilon} = \sum_{i=1}^{n-1} \max_{1 \le u \le p} \left[\sum_{j=1}^{u} t_{ij}^{\varepsilon,l} - \sum_{j=1}^{u-1} t_{i+1,j}^{\varepsilon,l} \right] + \sum_{j=1}^{p} t_{nj}^{\varepsilon,l}$, $l = \overline{1,k}$ – время выполнения l-й группы блоков всеми n процессами на p процессорах;

$\delta_l = \min\{\delta_l', \delta_l''\}$ – длина отрезка максимально возможного совмещения двух последовательных диаграмм Ганта по оси времени, δ_l' – разность между моментом начала выполнения j-го блока первым процессом для $(l+1)$-й группы блоков и моментом завершения выполнения j-го блока последним процессом для l-й группы блоков, δ_l'' – разность между началом выполнения первого блока i-м процессом для $(l+1)$-й группы блоков и моментом завершения выполнения p-го блока i-м процессом для l-й группы блоков, т. е.

$$\delta_l' = \min_{1 \le j \le p} \left[\sum_{w=j+1}^{p} t_{nw}^{\varepsilon,l} + \sum_{w=1}^{j-1} t_{1w}^{\varepsilon,l+1} \right],$$

$$\delta_l'' = \min_{1 \le i \le n} \left[T_l^{\varepsilon} + \sum_{q=1}^{i-1} \max_{1 \le u \le r} \left(\sum_{j=1}^{u} t_{qj}^{\varepsilon,l+1} - \sum_{j=1}^{u-1} t_{q+1,j}^{\varepsilon,l+1} \right) - \sum_{q=1}^{i-1} \max_{1 \le u \le p} \left(\sum_{j=1}^{u} t_{qj}^{\varepsilon,l} - \sum_{j=1}^{u-1} t_{q+1,j}^{\varepsilon,l} \right) - \sum_{j=1}^{p} t_{ij}^{\varepsilon,l} \right],$$

$$l = \overline{1,k-1}.$$

Здесь же рассмотрена процедура, которая позволяет решать задачу определения минимального общего времени выполнения *неоднородных* распределенных конкурирующих процессов $T_н^1(p,n,s,\varepsilon)$ с помощью математического аппарата *сетевых дуго-взвешенных* графов. Вершины в графе определяют моменты начала выполнения очередных p блоков i-го процесса в l-й группе блоков. Соединены вершины дугами трех типов: *горизонтальными* дугами, которые отражают продолжительность выполнения очередных p блоков i-го процесса в l-й группе блоков; *вертикальными* дугами, которые отражают времена задержек начала выполнения первого блока *(i+1)*-го процесса относительно начала выполнения первого блока i-го процесса в l-й группе блоков; *наклонными (диагональными)* дугами, отражающие времена задержек начала выполнения первого блока первого процесса в *(l+1)*-й группе блоков по отношению к первому блоку n-го процесса в l-й группе. Построенный таким образом сетевой дуго-взвешенный граф, полностью отображает во времени выполнение n неоднородных распределенных конкурирующих процессов на p процессорах в первом синхронном режиме с непрерывным выполнением s блоков структурированного программного ресурса каждым из процессов.

В 2.2 решены задачи нахождения минимального общего времени выполнения *однородных* распределенных конкурирующих процессов $T_o^1(p,n,s,\varepsilon)$ и *одинаково распределенных* $T_{op}^1(p,n,s,\varepsilon)$. Доказана следующая

Теорема 2.2. *Если взаимодействие процессов, процессоров и блоков подчинено условиям первого базового синхронного режима, то для любых* $p \geq 2$, $n \geq 2$, $s \geq 2$, $\varepsilon > 0$, *минимальное общее время выполнения одинаково распределенных конкурирующих процессов определяется по формулам:*

$$T_{op}^1(p,n,s,\varepsilon) = \begin{cases} \sum_{i=1}^{n} t_i^\varepsilon + (s-1)\left[t_n^\varepsilon + \sum_{i=2}^{n} \max\{t_{i-1}^\varepsilon - t_i^\varepsilon, 0\} \right], \ s \leq p, \\ kT_{op}^1(p,n,p,\varepsilon) - (k-1)\min\{\omega_1, \omega_2\}, \ s = kp, \ k > 1, \\ kT_{op}^1(p,n,p,\varepsilon) + T_{op}^1(p,n,r,\varepsilon) - (k-1)\min\{\omega_1,\omega_2\} - \min\{\xi_1,\xi_2\}, \\ s = kp + r, \ k \geq 1, \ 1 \leq r < p, \end{cases}$$

где величины $\omega_1 = (p-1)\min\{t_1^\varepsilon, t_n^\varepsilon\}$ *и* $\omega_2 = T_{op}^1(p,n,p,\varepsilon) - p\max\limits_{1 \leq i \leq n} t_i^\varepsilon$ *представляют собой временные отрезки максимально возможного совмещения* l*-й и* $(l+1)$*-й диаграмм Ганта, а* $\xi_1 = (r-1)\min\{t_1^\varepsilon, t_n^\varepsilon\} + (p-r)t_n^\varepsilon$ *и* $\xi_2 = T_{op}^1(p,n,p,\varepsilon) - \max\limits_{1 \leq i \leq n}\left(T_{op}^1(p,i,p,\varepsilon) - T_{op}^1(p,i,r,\varepsilon) + rt_i\right)$ *– временные* *отрезки максимально возможного совмещения по оси времени* k *-й и* $(k+1)$ *-й диаграмм.*

В 2.3 в рамках математической модели организации распределенной обработки конкурирующих процессов, введенной в 1.4, исследуется *второй синхронный* режим, т. е. когда для каждого блока структурированного программного ресурса момент завершения его выполнения для *i*-го процесса совпадает с моментом начала его выполнения для *(i+1)*-го процесса на том же процессоре. В условиях рассматриваемого режима для любых $p \geq 2$, $n \geq 2$, $s \geq 2$, $\varepsilon > 0$, получены временные характеристики реализации выполнения *неоднородных, однородных* и *одинаково распределенных* конкурирующих процессов. Остановимся на одинаково распределенных процессах, так как именно для этого класса в диссертации проведен сравнительный анализ и получены критерии эффективности и оптимальности.

Для определения минимального общего времени $T_{op}^2(p,n,s,\varepsilon)$ выполнения n конкурирующих *одинаково распределенных* процессов в многопроцессорной системе с p процессорами во *втором синхронном* режиме взаимодействия процессов, процессоров и блоков с учетом параметра $\varepsilon > 0$, имеет место

Т е о р е м а 2.3. *Минимальное общее время выполнения множества конкурирующих одинаково распределенных процессов во втором синхронном режиме при любых* $p \geq 2$, $n \geq 2$, $s \geq 2$, $\varepsilon > 0$, *определяется по формулам:*

$$T_{op}^2(p,n,s,\varepsilon) = \begin{cases} \sum_{i=1}^{n} t_i^{\varepsilon} + (s-1)\max\limits_{1\leq i\leq n} t_i^{\varepsilon}, \ s \leq p, \\ k\sum_{i=1}^{n} t_i^{\varepsilon} + (p-1)\max\limits_{1\leq i\leq n} t_i^{\varepsilon}, \ s = kp, \ k > 1, \\ (k+1)\sum_{i=1}^{n} t_i^{\varepsilon} + (r-1)\max\limits_{1\leq i\leq n} t_i^{\varepsilon}, \ s = kp + r, \ k \geq 1, \ 1 \leq r < p. \end{cases}$$

В 2.4 проведен *сравнительный анализ* асинхронного и двух синхронных режимов организации одинаково распределенных конкурирующих процессов с точки зрения временных затрат и с учетом параметра ε. В 2.1–2.3 для *асинхронного* и *второго синхронного* режимов получена формула:

$$T_{op}^{ac}(p,n,s,\varepsilon) = T_{op}^{2}(p,n,s,\varepsilon) = T_{\varepsilon}^{n} + (s-1)t_{\max}^{\varepsilon},$$

а для *первого синхронного* режима:

$$T_{op}^{1}(p,n,s,\varepsilon) = T_{\varepsilon}^{n} + (s-1)\left[t_{\varepsilon}^{n} + \sum_{i=2}^{n} \max\{t_{i-1}^{\varepsilon} - t_{i}^{\varepsilon}, 0\}\right],$$

где $T_{\varepsilon}^{n} = \sum_{i=1}^{n} t_{i}^{\varepsilon}$ – суммарное время выполнения каждого из блоков Q_j, $j=\overline{1,s}$, всеми n процессами с учетом накладных расходов ε, $t_{\max}^{\varepsilon} = \max\limits_{1\le i\le n} t_{i}^{\varepsilon}$, $t_{i}^{\varepsilon} = t_i + \varepsilon$.

В работе набор параметров $(t_1^{\varepsilon},\ t_2^{\varepsilon},\ \dots,\ t_n^{\varepsilon},\ T_{\varepsilon}^{n})$ одинаково распределенной системы конкурирующих процессов определен как *характеристический.*

Через $\beta = \left\{(t_1^{\varepsilon},\ t_2^{\varepsilon},\ \dots,\ t_n^{\varepsilon},\ T_{\varepsilon}^{n}) \;\middle|\; T_{\varepsilon}^{n} = \sum_{i=1}^{n} t_i^{\varepsilon},\ t_i^{\varepsilon} = t_i + \varepsilon > 0\right\}$ обозначено множество всех допустимых характеристических наборов систем одинаково распределенных конкурирующих процессов, из которого выделено подмножество характеристических наборов вида

$$\mathrm{H}(T_{\varepsilon}^{n}) = \{(t_1^{\varepsilon},\ t_2^{\varepsilon},\ \dots,\ t_n^{\varepsilon},\ T_{\varepsilon}^{n}) \in \beta \mid t_1^{\varepsilon} \le t_2^{\varepsilon} \le \dots \le t_l^{\varepsilon} \ge t_{l+1}^{\varepsilon} \ge \dots \ge t_n^{\varepsilon},\ l=\overline{1,n}\}.$$

Доказаны следующие утверждения:

Теорема 2.4. *Пусть* $\delta \in \mathrm{H}(T_{\varepsilon}^{n})$ *– характеристический набор любой одинаково распределенной системы с параметрами p, n, s и накладными расходами* $\varepsilon > 0$. *Тогда в случае* $2 \le s \le p$ *минимальные общие времена* T_{op}^{ac}, T_{op}^{1} *и* T_{op}^{2} *выполнения множества одинаково распределенных конкурирующих процессов в асинхронном и базовых синхронных режимах совпадают.*

Теорема 2.5. *Для любой одинаково распределенной системы с параметрами p, n, s и накладными расходами* $\varepsilon > 0$, *допустимый характеристический набор которой* $\delta \notin \mathrm{H}(T_{\varepsilon}^{n})$, *при* $2 \le s \le p$ *выполняются соотношения* $T_{op}^{1}(p,n,s,\varepsilon) > T_{op}^{ac}(p,n,s,\varepsilon) = T_{op}^{2}(p,n,s,\varepsilon)$.

В **третьей** главе получены необходимые и достаточные условия эффективности и оптимальности организации выполнения множества взаимодействующих одинаково распределенных конкурирующих процессов в условиях неограниченного и ограниченного параллелизма с учетом накладных расходов по времени их реализации.

В 3.1 в классе одинаково распределенных систем конкурирующих процессов выделен специальный подкласс *стационарных* систем, т. е. когда $t_1 = t_2 = ... = t_n = t$. Показано, что для всех трех базовых режимов в случае *стационарной одинаково распределенной* системы конкурирующих процессов в случае *неограниченного параллелизма* минимальное общее время их выполнения $\overline{T}_\varepsilon$ определяется равенством:

$$\overline{T}_\varepsilon(s \le p) = (n+s-1)t_\varepsilon \text{, где } t_\varepsilon = T^n / n + \varepsilon, \quad T^n = nt.$$

Одинаково распределенную систему конкурирующих процессов будем называть *эффективной* при фиксированных p, $s \ge 2$, если выполняется соотношение $\Delta_\varepsilon(n) = sT^n - T(p,n,s,\varepsilon) \ge 0$, где sT^n – время выполнения s блоков всеми n процессами в последовательном режиме. При наличии двух эффективных одинаково распределенных систем конкурирующих процессов будем считать, что первая более эффективна, чем вторая, если величина $\Delta_\varepsilon(n)$ первой системы не меньше соответствующей величины второй.

Для введенного подмножества одинаково распределенных систем доказано утверждение.

Теорема 3.1. *Для любой эффективной одинаково распределенной системы конкурирующих процессов при* $2 \le s \le p$ *и* $\varepsilon > 0$ *существует стационарная более эффективная одинаково распределенная система.*

Получены достаточное условие эффективности одинаково распределенной системы в случае неограниченного параллелизма и необходимое и достаточное условие существования эффективной системы одинаково распределенных конкурирующих процессов при достаточном числе процессоров в зависимости от величины накладных расходов.

Теорема 3.2. *Одинаково распределенная система конкурирующих процессов с параметрами* p, n, s, ε, *удовлетворяющая соотношениям* $3 \le s \le p$, $n = s \ne 3$, $sn \ge 2(n+s-1)$ *и* $0 < \varepsilon \le \min\limits_{1 \le i \le n} t_i$, *является эффективной.*

Теорема 3.3. *Для существования эффективной одинаково распределенной системы конкурирующих процессов с заданными пара-*

метрами $3 \le s \le p$, T^n, $\varepsilon > 0$ *необходимо и достаточно, чтобы выполнялись следующие условия:*

$$\varepsilon \le \begin{cases} \varphi(1+\sqrt{s}), & \text{если } \sqrt{s} - \text{целое,} \\ \max\{\varphi(1+[\sqrt{s}]),\ \varphi(2+[\sqrt{s}])\}, & \text{если } \sqrt{s} - \text{целое,} \end{cases}$$

где $\varphi(x) = \dfrac{(s-1)T^n(x-1)}{x(x+s-1)}$, $[x]$ – *наибольшее целое, не превосходящее* x.

В 3.2 получены необходимые и достаточные условия эффективности одинаково распределенных систем конкурирующих процессов в случае, когда число процессоров является *ограниченным*.

Теорема 3.4. *Если параметры одинаково распределенной системы* $n \ge 3$ *конкурирующих процессов в многопроцессорной системе с* p *процессорами удовлетворяют соотношениям* $s \ge 3$, $n = s \neq 3$ *и* $0 < \varepsilon \le \min\limits_{1 \le i \le n} t_i$, *то рассматриваемая система будет эффективной, если выполняются условия*

$$sn \ge \begin{cases} 2(kn+p-1), & \text{если } s = kp,\ k > 1, \\ 2((k+1)n+r-1), & \text{если } s = kp+r,\ k \ge 1,\ 1 \le r < p. \end{cases}$$

Теорема 3.5. *Для существования эффективной одинаково распределенной системы конкурирующих процессов с заданными параметрами* $p \ge 3$, T^n, $\varepsilon > 0$ *необходимо и достаточно, чтобы выполнялись следующие условия:*

1) при $s = kp$, $k > 1$,

$$\varepsilon \le \begin{cases} \varphi_1\left(\dfrac{1+\sqrt{p}}{k}\right), & \text{если } \dfrac{1+\sqrt{p}}{k} - \text{целое,} \\ \max\left\{\varphi_1\left(\left[\dfrac{1+\sqrt{p}}{k}\right]\right),\ \varphi_1\left(\left[\dfrac{1+\sqrt{p}}{k}\right]+1\right)\right\}, & \text{если } \dfrac{1+\sqrt{p}}{k} - \text{нецелое,} \end{cases}$$

где $\varphi_1(x) = (p-1)T^n(kx-1)/x(kx+p-1)$, *а* $[x]$ *– наибольшее целое, не превосходящее* x *;*

2) при $s = kp + r,\ k \geq 1,\ 1 \leq r < p$ *,*

$$\varepsilon \leq \begin{cases} \varphi_2(x), & \text{если } x \text{ – целое,} \\ \max\{\varphi_2([x]),\ \varphi_2([x]+1)\}, & \text{если } x \text{ – нецелое,} \end{cases}$$

где $\varphi_2(x) = \dfrac{\left[(p-1)kx + (r-1)(x-1)\right]\ T^n}{x\ \left[(k+1)x + r - 1\right]}$, $[x]$ *– наибольшее целое, не превосходящее* x*, где* $x = \dfrac{r-1}{(p-1)k+r-1}\left(1 + \sqrt{1 + \dfrac{(p-1)k+r-1}{k+1}}\right)$.

В 3.3 введено определение *оптимальной* одинаково распределенной системы: эффективная одинаково распределенная система называется *оптимальной*, если величина Δ_ε достигает наибольшего значения.

Решена задача оптимальности одинаково распределенной системы, состоящей из n конкурирующих процессов, для достаточного числа процессоров для всех трех базовых режимов.

Теорема 3.6. *Для того, чтобы эффективная одинаково распределенная система конкурирующих процессов была оптимальной при заданных* $2 \leq s \leq p$, T^n, $\varepsilon > 0$*, необходимо и достаточно, чтобы она была стационарной и число процессов* n_0 *в системе равнялось одному из чисел*

$$\left[\left[\sqrt{\frac{(s-1)T^n}{\varepsilon}}\right],\ \left[\sqrt{\frac{(s-1)T^n}{\varepsilon}}\right]+1\right] \cap [2, n],$$

в котором функция $\overline{\Delta}_\varepsilon(x)$ *достигает наибольшего значения. Здесь* $[x]$ *означает наибольшее целое, не превосходящее x, n – заданное число.*

Для решения задачи об оптимальности одинаково распределенной системы конкурирующих процессов в случае ограниченного параллелизма в асинхронном и втором синхронном режимах введены функции действительного аргумента *x* вида:

$$\overline{\Delta}_\varepsilon(x) = (s-1)T^n - \frac{(p-1)T^n}{x} - (kx+p-1)\varepsilon \text{, при } s = kp,\ k > 1,$$

$$\overline{\Delta}_\varepsilon(x) = (s-k-1)T^n - \frac{(r-1)T^n}{x} - \big((k+1)x+(r-1)\big)\varepsilon \text{, при}$$
$$s = kp+r,\ k \ge 1,\ 1 \le r < p.$$

Теорема 3.7. *Для того, чтобы эффективная одинаково распределенная система конкурирующих процессов в случае ограниченного параллелизма в асинхронном и втором синхронном режимах была оптимальной при заданных* $p \ge 2$, T^n, $\varepsilon > 0$, *необходимо и достаточно, чтобы она была стационарной и число процессов n в системе равнялось одному из чисел:*

1) $\left[\left[\sqrt{\dfrac{(p-1)T^n}{k\varepsilon}}\right], \left[\sqrt{\dfrac{(p-1)T^n}{k\varepsilon}}\right]+1\right]$, при $s = kp,\ k > 1$,

2) $\left[\left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right], \left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right]+1\right]$, при $s = kp+r,\ k \ge 1,\ 1 \le r < p$,

в котором функция $\overline{\Delta}_\varepsilon(x)$ *достигает наибольшего значения, где* $[x]$ *– наибольшее целое, не превосходящее* x.

В **четвертой** главе диссертации предложены приемы, с помощью которых можно достичь ускорения вычислений в многопроцессорных распределенных системах.

Основные результаты работы

1. Решены задачи нахождения минимального общего времени выполнения неоднородных, однородных и одинаково распределенных конкурирующих процессов в асинхронном и синхронных режимах с учетом дополнительных системных расходов, связанных с организацией параллельного использования блоков структурированного программного ресурса множеством конкурирующих процессов при распределенной обработке [1–2, 5, 8–10, 13–14].

2. Проведен сравнительный анализ времен выполнения в базовых режимах одинаково распределенных конкурирующих процессов при достаточном числе процессоров многопроцессорной системы [3, 5–6].

3. Для всех трех базовых режимов в случае достаточного числа процессоров, а для асинхронного и второго синхронного режимов взаимодействия конкурирующих процессов в общем случае, получены достаточные условия эффективности одинаково распределенных систем конкурирующих процессов [1, 3, 10, 15].

4. Определены условия существования эффективных систем одинаково распределенных конкурирующих процессов в зависимости от величины дополнительных системных расходов [1, 3, 10].

5. В условиях неограниченного и ограниченного параллелизма получены условия оптимальности одинаково распределенных систем конкурирующих процессов [4, 7, 11].

Список публикаций по теме диссертации

1. Капитонова Ю.В., Коваленко Н.С., Павлов П.А. Оптимальность систем одинаково распределенных конкурирующих процессов // Кибернетика и системный анализ. – 2005. – №6. – С. 3–10.
2. Коваленко Н.С., Павлов П.А. Эффективность и оптимальность структурирования программных ресурсов при распределенной обработке //.Труды минского института управления. – 2005. – №1. – С. 104–107.
3. Коваленко Н.С., Павлов П.А. Эффективность систем конкурирующих процессов с учетом накладных расходов // Доклады НАН Беларусі. Сер. фіз.-мат. навук. – 2005. – №6. – С. 32–36.
4. Коваленко Н.С., Павлов П.А. Системы одинаково распределенных конкурирующих процессов в условиях ограниченного параллелизма и их оптимальность // Доклады НАН Беларусі. Сер. фіз.–мат. навук. – 2006. – №2. – С. 25–29.
5. Павлов П.А. Анализ режимов организации одинаково распределенных конкурирующих процессов // Вестник БГУ. Сер. 1: Физ. Мат. Информ. – 2006. – №1. – С. 116–120.
6. Павлов П. А. Сравнительный анализ одинаково распределенных конкурирующих процессов с учетом дополнительных системных расходов // Вестник Фонда фундаментальных исследований. – 2006. – №1. – С. 55 – 58.
7. Коваленко Н.С., Павлов П.А. Оптимальность одинаково распределенных систем конкурирующих процессов // Информ. системы и технологии IST'2006: Материалы третьей Междунар. конф. Минск, 1–

3 ноября 2006 г. / Акад. упр. при Президенте Респ. Беларусь. – Минск, 2006. – С. 232–235.

8. Павлов П.А. Синхронный режим взаимодействия распределенных конкурирующих процессов с непрерывным выполнением блоков // Сицияльно-экономическое и гуманитарное развитие белорусского общества в XXI веке: Материалы респ. науч. конф. Минск, 16 дек. 2004 г. / БГЭУ. – Минск, 2005. – С. 426–427.

9. Павлов П.А. Коэффициенты ускорения и эффективности при распределенной обработке // Экономический механизм формирования национальной модели развития экономики РБ: Материалы науч.–практ. конф. Пинск, 22–23 февраля 2005 г. / БГЭУ. – Минск, 2005. – С. 98–100.

10. Павлов П.А. Эффективность систем одинаково распределенных конкурирующих процессов // Актуальные проблемы рыночной экономики: Материалы II науч.-практ. конф. Бобруйск, 7 апр. 2005 г. / БГЭУ. – Минск, 2005. – С. 141–144.

11. Павлов П.А. Оптимальность систем одинаково распределенных конкурирующих процессов при достаточном числе процессоров // Экономическое развитие Беларуси в контексте расширения Европейского Союза: Материалы Междунар. студ. науч. конф. Гродно, 12–13 мая 2005 г. В 2 ч. ч. 2 / ГрГУ. – Гродно, 2005. – С. 27–28.

12. Павлов П.А. Приемы ускорения вычислений // Исследования молодых ученых Пинщины: Материалы II науч.–практ. конф. Пинск, 14 мая 2005 г. / БГЭУ. – Пинск, 2005. – С. 106–110.

13. Павлов П.А. Минимальное общее время выполнения распределенных конкурирующих процессов в асинхронном режиме в случае неограниченного параллелизма // Механизм формирования социально–экономического развития регионов Республики Беларусь в условиях перехода к рыночной экономике: Материалы науч.-практ. конф. Пинск, 21–22 февраля 2006 г. / БГЭУ. – Минск, 2006. – С. 103–104.

14. Павлов П.А. Минимальное общее время выполнения неоднородных распределенных конкурирующих процессов в первом синхронном режиме // Механизмы устойчивого развития инновационных социально–экономических систем: Материалы международной науч.–практ. конф. Бобруйск, 30 марта 2006 г. / БГЭУ. – Минск, 2006. – С. 162–164.

15. Павлов П.А. Критерии эффективности организации выполнения множества распределенных конкурирующих процессов // Социально–экономическое и историко-культурное развитие Полесского региона в XXI веке: Материалы международной науч.–практ. конф. Пинск, 5–6 мая 2006 г. / КУП. – Пинск, 2006. – С. 163–165.