ISSN 1819-3161

ПРОБЛЕМЫ УПРАВЛЕНИЯ

3/2010

CONTROL SCIENCES

CONTROL SCIENCES

Научно-технический журнал
6 номеров в год
ISSN 1819-3161

УЧРЕДИТЕЛЬ
Учреждение Российской академии наук
Институт проблем управления им. В.А. Трапезникова РАН

Издатель
ООО «СенСиДат-Контрол»

Адрес редакции
117997, ГСП-7, Москва,
ул. Профсоюзная, д. 65, к. 410.
Тел./факс (495) 334-92-00

E-mail: pu@ipu.ru

Интернет: http://pu.mtas.ru

Оригинал-макет и электронная версия подготовлены ООО «Адвансед Солюшнз»

Отпечатано в ИПУ РАН

Заказ № 80

Подписано в печать 8.06.2010 г.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций

Свидетельство о регистрации ПИ №77-11963 от 06 марта 2002 г.

Журнал входит в Перечень ведущих рецензируемых журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук

Подписные индексы:
80508 и **81708** в каталоге Роспечати
38006 в объединенном каталоге «Пресса России»

Цена свободная

ПРОБЛЕМЫ УПРАВЛЕНИЯ

3.2010

СОДЕРЖАНИЕ

УДК 681.3.06

ОРГАНИЗАЦИЯ ОДНОРОДНЫХ КОНКУРИРУЮЩИХ ПРОЦЕССОВ ПРИ РАСПРЕДЕЛЕННОЙ КОНВЕЙЕРНОЙ ОБРАБОТКЕ

П.А. Павлов

Получены формулы и оценки минимального общего времени выполнения однородных распределенных конкурирующих процессов. Проведен сравнительный анализ режимов взаимодействия процессов, процессоров и блоков структурированного программного ресурса с учетом дополнительных системных расходов.

Ключевые слова: распределенный процесс, программный ресурс, однородная система, асинхронный режим, синхронный режим, структурирование, параллелизм.

ВВЕДЕНИЕ

В различных областях человеческой деятельности постоянно приходится сталкиваться с большими задачами, эффективное решение которых связано с распараллеливанием процессов вычислений. В связи с этим особую актуальность приобретают задачи построения и исследования математических моделей обработки распараллеливаемых программных ресурсов с последующей их конвейеризацией по процессам и процессорам. Практический интерес представляют задачи определения временных характеристик однородных конкурирующих процессов, сравнительного анализа режимов взаимодействия распределенных процессов, получения условий эффективности систем однородных конкурирующих процессов с учетом накладных расходов по времени их реализации. Именно этим задачам посвящена настоящая статья.

1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ КОНКУРИРУЮЩИХ ПРОЦЕССОВ

В качестве конструктивных элементов для построения математических моделей распределенных вычислительных систем служат понятия процесса и программного ресурса.

Как и в работе [1], *процесс* будем рассматривать как последовательность блоков (команд, процедур) Q_1, Q_2, ..., Q_s, для выполнения которых используется множество процессоров (процессорных узлов, обрабатывающих устройств). Процесс называется *распределенным*, если все блоки или часть из них обрабатываются разными процессорами. Для ускорения выполнения процессы могут обрабатываться параллельно, взаимодействуя путем обмена информацией. Такие процессы называются *кооперативными* или *взаимодействующими* процессами [2].

Понятие *ресурса* употребляется для обозначения любых объектов вычислительной системы, которые могут быть использованы процессами для своего выполнения. *Реентерабельные* (многократно используемые) ресурсы характеризуются возможностью их одновременного использования несколькими процессами. Для параллельных систем характерна ситуация, когда одну и ту же последовательность блоков или ее часть необходимо процессорам выполнять многократно, такую последовательность будем называть *программным ресурсом* (ПР), а множество соответствующих процессов — *конкурирующими*.

Решая проблему распределения программных ресурсов, неявно решаем задачи эффективного использования ресурсов остальных категорий. В свою очередь, эффективно решая задачу распределения программных ресурсов, решаем проблему сокращения времени реализации множества параллельных распределенных процессов. Данные задачи можно эффективно решать, опираясь на методы структурирования, принципы распараллеливания и конвейеризации [2].

Как и в работах [1—5], математическая модель распределенной обработки конкурирующих процессов включает в себя p процессоров многопроцессорной системы (МС), n конкурирующих процессов, s блоков Q_1, Q_2, ..., Q_s структурированного на блоки программного процесса, матрицу $T_p = [t_{ij}]$ времен выполнения j-х блоков i-ми конкурирующими процессами, причем $p \geqslant 2$, $n \geqslant 2$, $s \geqslant 2$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant s$. Будем предполагать, что все n процессов используют одну копию структурированного на блоки ПР, а на множестве блоков установлен линейный порядок их выполнения.

Введем в рассмотрение параметр $\tau > 0$, характеризующий время (системные расходы), затрачиваемые МС на организацию параллельного выполнения блоков программного ресурса множеством распределенных конкурирующих процессов. В дальнейшем будем говорить, что перечисленные объекты математической модели образуют *систему* распределенных конкурирующих процессов.

Определение 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 2. Система n распределенных конкурирующих процессов называется *однородной*, если времена выполнения j-го блока каждым из i-х процессов равны, т. е. $t_{ij} = t_j$, $i = \overline{1, n}$, $j = \overline{1, s}$. ♦

В работах [1, 2] введены режимы взаимодействия процессов, процессоров и блоков структурированного программного ресурса.

Асинхронный режим взаимодействия процессоров, процессов и блоков предполагает отсутствие простоев процессоров при условии готовности блоков, а также невыполнение блоков при наличии процессоров. На рис. 1 представлен пример диаграммы Ганта, которая отображает выполнение $n = 3$ однородных распределенных конкурирующих процессов, использующих структурированный на $s = 8$ блоков ПР в МС с $p = 3$ процессорами. Матрица времен выполнения блоков n конкурирующими процессами с учетом параметра τ имеет вид

$$T_p^{\tau} = \begin{bmatrix} 2 & 4 & 2 & 3 & 1 & 4 & 2 & 3 \\ 2 & 4 & 2 & 3 & 1 & 4 & 2 & 3 \\ 2 & 4 & 2 & 3 & 1 & 4 & 2 & 3 \end{bmatrix}.$$

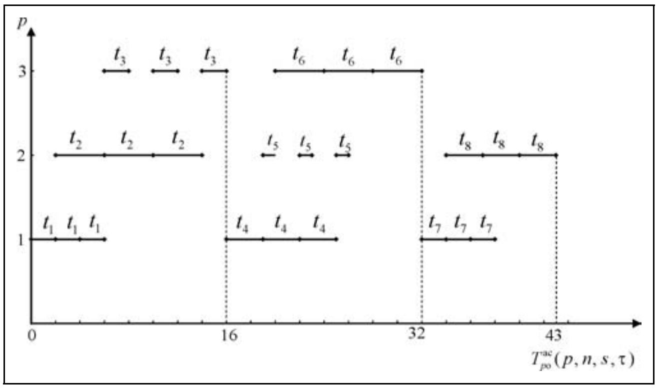

Рис. 1. Диаграмма асинхронного режима

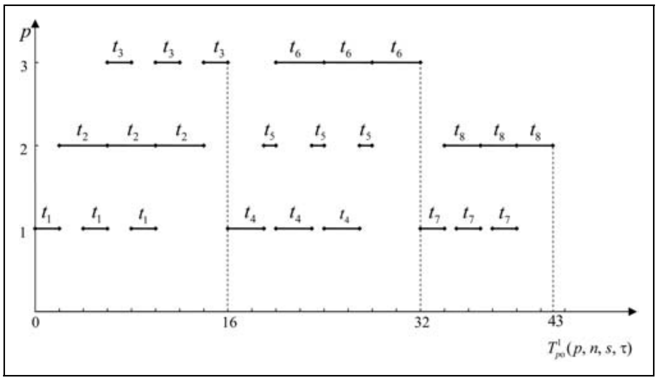

Рис. 2. Диаграмма первого синхронного режима

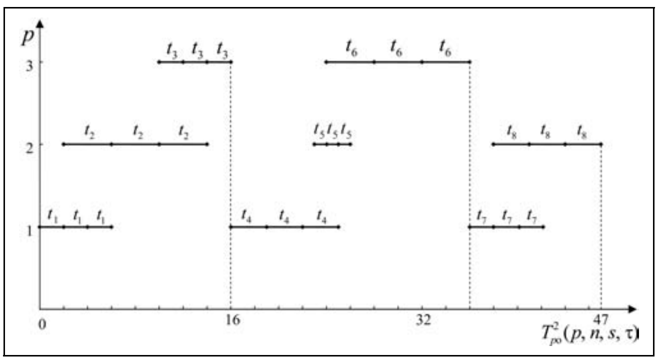

Рис. 3. Диаграмма второго синхронного режима

Так как $s > p$, $s = kp + r$, $k = 2$, $r = 2$, то мы имеем случай *ограниченного* параллелизма [1]. В этом случае исходную матрицу времен выполнения блоков T_p^{τ} разбиваем на $k + 1$ подматрицу T_l^{τ}, $l = \overline{1, k+1}$, размерностью $n \times p$ каждая, за исключением последней T_{k+1}^{τ}, которая будет содержать только r столбцов, а остальные $p - r$ столбцов будут нулевыми. По каждой из подматриц T_l^τ, $l = \overline{1, k+1}$, строим $k + 1$ линейную диаграмму Ганта, каждая из которых отображает во времени выполнение p блоков структурированного ПР на p процессорах всеми n процессами. Последняя диаграмма отображает выполнение последних r блоков всеми n процессами.

Первый синхронный режим обеспечивает непрерывное выполнение блоков программного ресурса внутри каждого из процессов (рис. 2).

Второй синхронный режим обеспечивает непрерывное выполнение каждого блока всеми процессами (рис. 3).

2. ВРЕМЯ РЕАЛИЗАЦИИ ОДНОРОДНЫХ СИСТЕМ РАСПРЕДЕЛЕННЫХ КОНКУРИРУЮЩИХ ПРОЦЕССОВ

В работах [1, 2] исследованы базовые асинхронный и синхронные режимы, возникающие при организации распределенных процессов в условиях конкуренции за общий программный ресурс. В рамках этих режимов получены математические соотношения для вычисления значений минимального общего времени выполнения неоднородных распределенных конкурирующих процессов в случаях *неограниченного* ($s \leqslant p$) и *ограниченного* ($s > p$) параллелизма по числу процессоров многопроцессорной системы.

Рассмотрим однородную систему распределенных конкурирующих процессов. Пусть t_1^τ, t_2^τ, ..., t_s^τ — длительности выполнения каждого блока структурированного программного ресурса с учетом параметра $\tau > 0$.

При $s \leqslant p$ для вычисления минимального общего времени в *асинхронном* режиме $T_{po}^{ac}(p, n, s, \tau)$ и *первом синхронном* режиме $T_{po}^{1}(p, n, s, \tau)$ получим, что $T_{po}^{ac.1}(p, n, s, \tau) = \sum\limits_{j=1}^{s} t_j^\tau + (n-1)\max\limits_{1 \le j \le s} t_j^\tau$.

Далее рассмотрим случай, когда $s = kp$, $k > 1$. Введем следующие обозначения: $t_j^{\tau, l} = t_{(l-1)p+j} + \tau$ — время выполнения j-го блока l-й группы всеми n процессами, $j = \overline{1, p}$, $l = \overline{1, k}$; $T_l = \sum\limits_{j=1}^{p} t_j^{\tau, l} + (n-1) \times \max\limits_{1 \le j \le p} t_j^{\tau, l}$ — общее время выполнения l-й группы блоков всеми n процессами на p процессорах, $l = \overline{1, k}$; $E_l^j = \sum\limits_{w=1}^{j} t_w^{\tau, l} + (n-1)\max\limits_{1 \le w \le p} t_w^{\tau, l}$ — время завершения выполнения $[(l-1)p + j]$-го блока программного ресурса всеми n процессами на j-м процессоре, $j = \overline{1, p}$, $l = \overline{1, k}$. Тогда общее время выполнения n конкурирующих распределенных однородных процессов в случае $s = kp$, $k > 1$, будет определяется как сумма длин составляющих диаграмм Ганта с учетом максимально допустимого совмещения по оси времени, т. е.

$$T_{po}^{ac.1}(p, n, s = kp, \tau) = \sum_{l=1}^{k} T_l - \sum_{l=1}^{k-1} \min\{\varphi'_l, \varphi''_l\}.$$

Здесь φ'_l — отрезок возможного совмещения по оси времени, представляющий собой 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, а φ''_l — представляет собой 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, которые вычисляются по формулам:

$$\varphi'_l = \min_{1 \le j \le p}\left[T_l + \sum_{w=1}^{j-1} t_w^{\tau, l+1} - E_l^j\right] = \min_{1 \le j \le p}\left[\sum_{w=j+1}^{p} t_w^{\tau, l} + \sum_{w=1}^{j-1} t_w^{\tau, l+1}\right],$$

$$\varphi''_l = (n-1)\min\left[\max_{1 \le j \le p} t_j^{\tau, l},\ \max_{1 \le j \le p} t_j^{\tau, l+1}\right],$$
$$l = \overline{1, k-1}.$$

В случае $s = kp + r$, $k \geqslant 1$, $1 \leqslant r \leqslant p$, минимальное общее время в рассматриваемых режимах определяется по формуле:

$$T_{po}^{ac.1}(p, n, s = kp + r, \tau) = \sum_{l=1}^{k} T_l + T_{k+1} - \sum_{l=1}^{k-1}\min\{\varphi'_l, \varphi''_l\} - \min\{\varphi'_k, \varphi''_k\},$$

где $T_{k+1} = \sum\limits_{j=1}^{r} t_j^{\tau, k+1} + (n-1)\max\limits_{1 \le j \le r} t_j^{\tau, k+1}$ — время выполнения $(k + 1)$-й группы r блоков всеми n процессами, $\varphi'_k = \min\limits_{1 \le j \le r}\left[\sum\limits_{w=j+1}^{p} t_w^{\tau, k} + \sum\limits_{w=1}^{j-1} t_w^{\tau, k+1}\right]$ — разность между моментом начала выполнения j-го

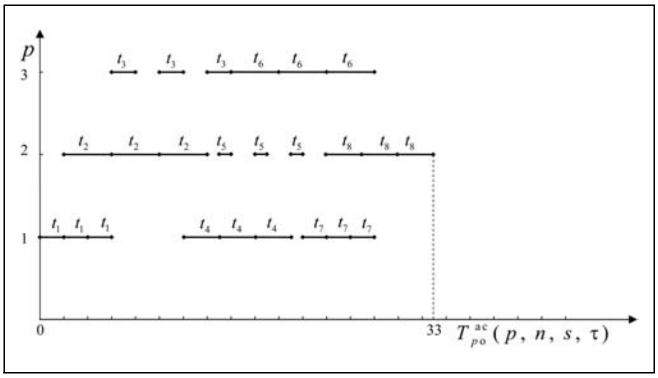

Рис. 4. Совмещенная диаграмма Ганта (асинхронный режим)

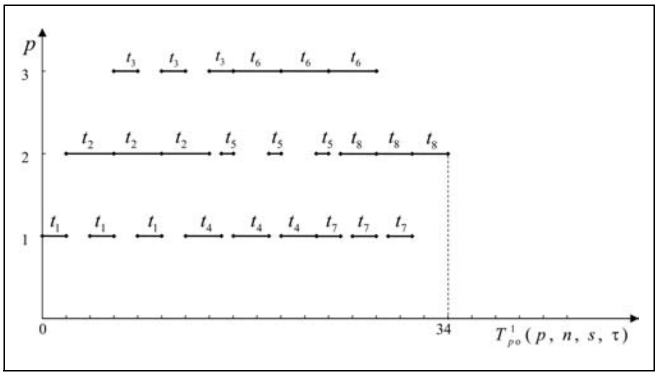

Рис. 5. Совмещенная диаграмма Ганта (первый синхронный режим)

блока первым процессом для $(k + 1)$-й группы блоков и моментом завершения выполнения j-го блока последним процессом для k-й группы блоков, $\varphi''_k = (n - 1)\min[\max\limits_{1 \le j \le p} t_j^{\tau, l}, \max\limits_{1 \le j \le r} t_j^{\tau, l+1}]$ — разность между началом выполнения первого блока i-м процессом для $(k + 1)$-й группы блоков и моментом завершения выполнения p-го блока i-м процессом для k-й группы блоков.

Из рис. 1 и 2 видно, что $T^{\text{ac}}_{\text{po}}(p, n, s, \tau) = T^{1}_{po}(p, n, s, \tau) = 43$. Однако это время можно существенно сократить, если воспользоваться приемом совмещения последовательных диаграмм Ганта по оси времени справа налево на размер максимально допустимого совмещения (рис. 4 и 5).

Если взаимодействие процессов, процессоров и блоков осуществляется во *втором синхронном* режиме (см. рис. 3), при котором 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, $i = \overline{1, n-1}$, то минимальное общее время $T^{2}_{po}(p, n, s, \tau)$ выполнения n однородных процессов на p процессорах определяется по формулам:

$$T^{2}_{po}(p, n, s, \tau) = \sum_{j=1}^{s} t_j^{\tau} + (n - 1) \times \left[t_s^{\tau} + \sum_{j=2}^{s} \max\{t_{j-1}^{\tau} - t_j^{\tau}, 0\} \right] \text{ при } s \leqslant p,$$

$$T^{2}_{po}(p, n, s, \tau) \leqslant$$

$$\leqslant \begin{cases} \sum\limits_{l=1}^{k} T_l - \sum\limits_{l=1}^{k-1} \min\{\psi'_l, \psi''_l\} \text{ при } s = kp, k > 1; \\ \sum\limits_{l=1}^{k} T_l + T_{k+1} - \sum\limits_{l=1}^{k-1} \min\{\psi'_l, \psi''_l\} - \min\{\psi'_l, \psi''_l\} \\ \text{при } s = kp + r, k \ge 1, 1 \le r < p. \end{cases} \quad (1)$$

Значения T_l, ψ'_l, ψ''_l, T_{k+1}, ψ'_k, ψ''_k вычисляются соответственно по формулам:

- $T_l = \sum\limits_{j=1}^{p} t_j^{\tau, l} + (n-1)\left[t_p^{\tau, l} + \sum\limits_{j=2}^{p} \max\{t_{j-1}^{\tau, l} - t_j^{\tau, l}, 0\}\right]$ — общее время выполнения l-х p блоков программного ресурса всеми n процессами на p процессорах, $l = \overline{1, k}$;
- ψ'_l и ψ''_l — отрезки возможного совмещения двух последовательных диаграмм по оси времени:

$$\psi'_l = \min_{1 \le j \le p}\{T_l + E_j^{l+1} - nt_j^{\tau, l+1} - E_j^l\},$$

$$\psi''_l = (n - 1)\min\{t_1^{\tau, l+1}, t_p^l\}, \quad l = \overline{1, k-1};$$

- $E_j^l = \sum\limits_{w=1}^{j} t_w^{\tau, l} + (n - 1)\left[t_j^{\tau, l} + \sum\limits_{w=2}^{j} \max\{t_{w-1}^{\tau, l} - t_w^{\tau, l}, 0\}\right]$, $j = \overline{1, p}$, $l = \overline{1, k}$ — время завершения выполнения $[(l - 1)p + j]$-го блока программного ресурса всеми n процессами на j-м процессоре;
- $T_{k+1} = \sum\limits_{j=1}^{r} t_j^{\tau, k+1} + (n - 1)\left[t_r^{\tau, k+1} + \sum\limits_{j=2}^{r} \max\{t_{j-1}^{\tau, k+1} - t_j^{\tau, k+1}, 0\}\right]$ — время выполнения $(k + 1)$-х r блоков для всех n процессов;

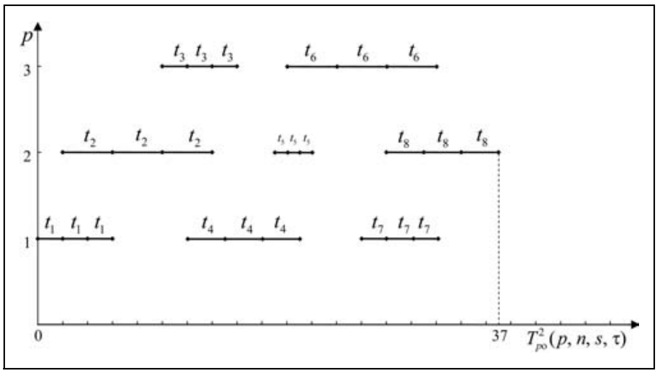

Рис. 6. Совмещенная диаграмма Ганта (второй синхронный режим)

- $\min\{\psi'_k,\ \psi''_k\}$ — максимальное совмещение по оси времени k-й и $(k + 1)$-й диаграмм:

$$\psi'_k = \min_{1 \leq j \leq r}\{T_k + E_j^{k+1} - n t_j^{\tau, k+1} - E_j^k\},$$
$$\psi''_k = (n - 1)\min\{t_1^{\tau, k+1},\ t_p^{\tau, k}\}.$$

В формуле (1) стоит знак неравенства, так как каждое значение $\min\{\psi'_l,\ \psi''_l\}$, $l = \overline{1, k-1}$, учитывает только максимально допустимое совмещение по оси времени между парами соседних диаграмм Ганта, но не всегда учитывает возможные совмещения между подряд идущими группами блоков, выполняющихся на одном и том же процессоре в двух соседних диаграммах.

С учетом максимально допустимого суммарного совмещения диаграмма Ганта для второго синхронного режима будет иметь вид, представленный на рис. 6.

3. АНАЛИЗ РЕЖИМОВ ОРГАНИЗАЦИИ РАСПРЕДЕЛЕННЫХ КОНКУРИРУЮЩИХ ПРОЦЕССОВ

Определенный теоретический и практический интерес представляет задача сравнительного анализа соотношений для определения минимального общего времени выполнения множества распределенных конкурирующих процессов. Проведем такой анализ для класса однородных систем с учетом дополнительных системных расходов $\tau > 0$.

Рассмотрим однородную систему распределенных конкурирующих процессов с временами выполнения блоков структурированного на блоки программного процесса $t_1^\tau,\ t_2^\tau,\ ...,\ t_s^\tau$. Обозначим через $T_p^\tau = \sum_{j=1}^{s} t_j^\tau$ суммарное время выполнения программного ресурса каждым из процессов с учетом накладных (системных) расходов и назовем набор параметров $(t_1^\tau,\ t_2^\tau,\ ...,\ t_s^\tau,\ T_p^\tau)$ данной системы *характеристическим.*

Пусть $\beta = \left\{(t_1^\tau,\ t_2^\tau,\ ...,\ t_n^\tau,\ T_p^\tau) \mid T_p^\tau = \sum_{j=1}^{s} t_j^\tau,\ t_j^\tau = t_j + \tau > 0, j = \overline{1, s}\right\}$ — множество всех допустимых характеристических наборов систем однородных конкурирующих процессов. Выделим из множества β подмножество характеристических наборов вида:

$$\mathrm{H}(T_p^\tau) = \{(t_1^\tau,\ t_2^\tau,\ ...,\ t_s^\tau,\ T_p^\tau) \in \beta \mid t_1^\tau \leqslant t_2^\tau \leqslant ... \leqslant t_l^\tau \geqslant t_{l+1}^\tau \geqslant ... \geqslant t_s^\tau,\ l = \overline{1, s}\}.$$

Тогда для введенного подмножества характеристических наборов справедливы следующие утверждения.

Теорема 1. *Пусть* $\delta \in \mathrm{H}(T_p^\tau)$ *— характеристический набор любой однородной системы с параметрами* p, n, $s \geqslant 2$ *и накладными расходами* $\tau > 0$. *Тогда в случае неограниченного параллелизма минимальные общие времена* $T_{\text{ро}}^{\text{ас}}$, $T_{\text{ро}}^1$ *и* $T_{\text{ро}}^2$ *выполнения множества однородных распределенных конкурирующих процессов в асинхронном и базовых синхронных режимах совпадают.* ♦

Пусть $t_l^\tau = \max_{1 \leq j \leq s} t_j^\tau$. Тогда для асинхронного и первого синхронного режима с непрерывным переходом по блокам для любого характеристического допустимого набора однородной системы, в том числе и для любого характеристического набора $\delta \in \mathrm{H}(T_p^\tau)$ при $2 \leqslant s \leqslant p$, имеют место равенства:

$$T_{\text{ро}}^{\text{ас}}(p, n, s, \tau) = T_{\text{ро}}^1(p, n, s, \tau) = T_p^\tau + (n - 1)t_l^\tau,$$

где $T_p^\tau = \sum_{j=1}^{s} t_j^\tau$, $t_j^\tau = t_j + \tau$, $j = \overline{1, s}$.

Если взаимодействие процессов, процессоров и блоков осуществляется во втором синхронном режиме (с непрерывным переходом по процессам), то для любого характеристического набора из множества β при $2 \leqslant s \leqslant p$ выполняется равенство:

$$T_{\text{ро}}^2(p, n, s, \tau) = \sum_{j=1}^{s} t_j^\tau + (n - 1) \times$$
$$\times \left[t_s^\tau + \sum_{j=2}^{s}\max\{t_{j-1}^\tau - t_j^\tau,\ 0\}\right].$$

Теорема 1 будет доказана, если будет выполняться равенство $t_s^\tau + \sum_{j=2}^{s}\max\{t_{j-1}^\tau - t_j^\tau,\ 0\} = t_l^\tau$. Учитывая, что $t_l^\tau = \max_{1 \le j < s} t_j^\tau$, для всех номеров $j \leqslant l$ имеет место равенство $\sum_{j=2}^{l} \max\{t_{j-1}^\tau - t_j^\tau, 0\} = 0$, а для $j > l$ имеет место $\sum_{j=l+1}^{s} \max\{t_{j-1}^\tau - t_j^\tau, 0\} = t_l^\tau - t_s^\tau$. Следовательно, $t_s^\tau + \sum_{j=2}^{s} \max\{t_{j-1}^\tau - t_j^\tau, 0\} = t_s^\tau + t_l^\tau - t_s^\tau = t_l^\tau$, что и требовалось доказать.

Теорема 2. *Для любой однородной распределенной системы с параметрами p, n, s и накладными расходами* $\tau > 0$, *допустимый характеристический набор которой* $\delta \notin \mathrm{H}(T_p^\tau)$, *при* $2 \leqslant s \leqslant p$ *выполняются соотношения:* $T_{\text{ро}}^2(p, n, s, \tau) > T_{\text{ро}}^{\text{ас}}(p, n, s, \tau) = T_{\text{ро}}^1(p, n, s, \tau)$. ♦

Условие теоремы 2 равносильно неравенству $t_s^\varepsilon + \sum_{j=2}^{s} \max\{t_{j-1}^\tau - t_j^\tau, 0\} - \max_{1 \le j \le s} t_j^\tau > 0$. Доказательство последнего проводится индукцией по числу блоков s, $s \geqslant 2$.

4. ЭФФЕКТИВНОСТЬ СИСТЕМ ОДНОРОДНЫХ КОНКУРИРУЮЩИХ ПРОЦЕССОВ В УСЛОВИЯХ НЕОГРАНИЧЕННОГО ПАРАЛЛЕЛИЗМА

Введем следующее определение, которое выделяет в классе однородных систем конкурирующих процессов специальный подкласс так называемых *равномерных* систем.

Определение 1. Однородную распределенную систему конкурирующих процессов назовем *равномерной*, если выполняется цепочка равенств $t_1^\tau = t_2^\tau = \ldots = t_s^\tau = t^\tau$. ♦

Согласно теореме 1 для однородных систем конкурирующих процессов минимальное общее время с учетом накладных расходов $\tau > 0$ для всех трех базовых режимов, указанных в § 1, в случае $s \leqslant p$ вычисляется по формуле:

$$T_{\text{ро}}^{\text{ас}.1,2}(p, n, s, \tau) = T_p^\tau + (n-1)t_{\max}^\tau, \qquad (1)$$

где $T_p^\tau = \sum_{j=1}^{s} t_j^\tau$, $t_j^\tau = t_j + \tau, j = \overline{1, s}$, $t_{\max}^\tau = \max_{1 \le j \le s} t_j^\tau$.

В случае равномерной однородной системы конкурирующих процессов минимальное общее время их выполнения определяется равенством

$$\overline{T}(p, n, s, \tau) = (n + s - 1)t^\tau, \qquad (2)$$

где $t^\tau = T^s/s + \tau$, $T^s = st$.

Определение 4. Однородную систему распределенных конкурирующих процессов будем называть *эффективной* при фиксированных p, $n \geqslant 2$, если выполняется соотношение $\Delta_\tau(s) = nT^s - \overline{T}(p, n, s, \tau) \geqslant 0$, где nT^s — время выполнения n процессов в последовательном режиме, а $T^s = \sum_{j=1}^{s} t_j$. ♦

При наличии двух эффективных однородных систем конкурирующих процессов будем считать, что первая более эффективна, чем вторая, если величина $\Delta_\tau(s)$ первой системы не меньше соответствующей величины второй. Для введенного подмножества однородных систем справедливо следующее утверждение.

Теорема 3. *Для любой эффективной однородной системы конкурирующих процессов при* $s \leqslant p$ *и* $\tau > 0$ *существует более эффективная равномерная однородная распределенная система.* ♦

Рассмотрим любую эффективную однородную распределенную систему. Согласно определению 4, условие ее эффективности с учетом формулы (1) записывается в виде следующего неравенства:

$$\Delta_\tau(s \leqslant p) = (n-1)(T^s - t_{\max}^s) - (n + s - 1)\tau \geqslant 0, \qquad (3)$$

где $T^s = \sum_{j=1}^{s} t_j$, $t_{\max}^s = \max_{1 \le j \le s} t_j$.

Для любой равномерной однородной распределенной системы с учетом равенства (2) имеем:

$$\overline{\Delta}_\tau(s \leqslant p) = (n-1)(T^s - t) - (n + s - 1)\tau \geqslant 0,$$

где $t = T^s/s$.

Справедливость теоремы 3 доказывает выполнение неравенства $\overline{\Delta}_\tau \geqslant \Delta_\tau$ для введенных эффективных систем.

Следующее утверждение ставит достаточное условие эффективности однородной системы в случае неограниченного параллелизма.

Теорема 4. *Однородная система конкурирующих процессов с параметрами* p, n, s, τ, *удовлетворяющая соотношениям* $3 \leqslant s \leqslant p$, $s = n \ne 3$, $ns \geqslant 2(n + s - 1)$ *и* $0 < \tau \leqslant \min_{1 \le j \le s} t_j$, *является эффективной.* ♦

Согласно неравенству (3) условие эффективности равносильно неравенству

$$\frac{T^s - t_{\max}^s}{\tau} \geqslant \frac{n + s - 1}{n - 1}. \qquad (4)$$

Следовательно, для доказательства теоремы 4 достаточно убедиться в справедливости неравенства (4). Непосредственная проверка показывает, что следствием соотношений $0 < \tau \leqslant \min\limits_{1 \le j \le s} t_j = t^s_{\min}$ является цепочка неравенств:

$$\frac{T^s - t^s_{\max}}{\tau} \geqslant \frac{(s-1)t^s_{\min}}{\tau} \geqslant s - 1, \qquad (5)$$

так как в силу выбора параметра τ выполняется неравенство $t^s_{\min}/\tau \geqslant 1$.

Из $ns \geqslant 2(n + s - 1)$ следует справедливость неравенства

$$s - 1 \geqslant \frac{n+s-1}{n-1}. \qquad (6)$$

Проверка показывает, что неравенство (4) является следствием неравенств (5) и (6).

Далее формулируется критерий существования эффективной однородной системы распределенных конкурирующих процессов при достаточном числе процессоров в зависимости от накладных расходов τ.

Теорема 5. *Для существования эффективного структурирования программного ресурса при заданных параметрах* $3 \leqslant s \leqslant p$, T^s, $\tau > 0$ *необходимо и достаточно, чтобы выполнялись следующие условия:*

$$\tau \leqslant \begin{cases} \varphi(1+\sqrt{n}), \text{ если } \sqrt{n} \text{ — целое}, \\ \max\{\varphi(1+[\sqrt{n}]), \varphi(2+[\sqrt{n}])\}, \\ \qquad \text{если } \sqrt{n} \text{ — не целое}, \end{cases}$$

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

Условие эффективности любой однородной распределенной системы конкурирующих процессов равносильно неравенству:

$$\varepsilon \leqslant \frac{(n-1)T^s(s-1)}{s(n+s-1)}. \qquad (7)$$

Введем в рассмотрение функцию $\varphi(x) = \dfrac{(n-1)T^s(x-1)}{x(n+x-1)}$.

Нетрудно проверить, что она достигает своего максимума при $x > 0$ в точке $x = 1 + \sqrt{n}$. Выбрав в качестве эффективного структурирование на s блоков, при котором $s = x = 1 + \sqrt{n}$, если $\sqrt{n}$ — целое, или $s = x \in \{1 + [\sqrt{n}], 2 + [\sqrt{n}]\}$, если $\sqrt{n}$ — не целое, докажем необходимость.

Достаточность следует из неравенства (7), поскольку функция $\varphi(x)$ достигает наибольшего значения при $x = 1 + \sqrt{n}$, если $\sqrt{n}$ — целое, или $x \in \{1 + [\sqrt{n}], 2 + [\sqrt{n}]\}$, если $\sqrt{n}$ — не целое.

ЗАКЛЮЧЕНИЕ

В работе получены оценки минимального общего времени реализации однородных конкурирующих процессов, проведен анализ режимов взаимодействия однородных процессов, процессоров и блоков структурированного программного ресурса, показано, что эффективную однородную систему следует искать среди равномерных однородных систем, сформулированы критерии эффективности однородных систем в условиях неограниченного параллелизма и с учетом накладных расходов. Полученные формулы служат основой для решения задач оптимизации числа блоков при заданных остальных параметрах многопроцессорной системы, нахождения оптимального числа процессоров при заданных объемах вычислений и (или) директивных сроках реализации процессов, исследования всевозможных смешанных режимов организации выполнения параллельных процессов при распределенной обработке, в том числе с учетом ограниченного числа копий структурированного программного ресурса и др.

ЛИТЕРАТУРА

1. *Иванников В.П., Коваленко Н.С., Метельский В.М.* О минимальном времени реализации распределенных конкурирующих процессов в синхронных режимах // Программирование. — 2000. — № 5. — С. 44—52.
2. *Коваленко Н.С., Самаль С.А.* Вычислительные методы реализации интеллектуальных моделей сложных систем. — Минск: Бел. навука, 2004. — 166 с.
3. *Kapitonova Yu.V., Kovalenko N.S., Pavlov P.A.* Optimality of systems of identically distributed competing processes // Cybernetics and Systems Analysis. — New York: Springer, — 2006. — P. 793—799.
4. *Павлов П.А.* Анализ режимов организации одинаково распределенных конкурирующих процессов // Вестник БГУ / Сер. 1: Физ.-мат. информ. — 2006. — № 1. — С. 116—120.
5. *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.

Статья представлена к публикации членом редколлегии чл.-корр. РАН П.П. Пархоменко.

Павлов Павел Александрович — канд. физ.-мат. наук, доцент, Полесский государственный университет, Республика Беларусь, г. Пинск, ☎+375 (0) 165-32-21-23, ✉ p.a.pavlov@mail.ru.