УДК 681.3.06:519

П.А. ПАВЛОВ

ОПТИМАЛЬНОСТЬ СИСТЕМ ОДИНАКОВО РАСПРЕДЕЛЕННЫХ КОНКУРИРУЮЩИХ ПРОЦЕССОВ

Necessary conditions and criteria of optimality of systems of equally distributed competing processes are received in view of an overhead charge on time of their realization.

Проблема оптимального распределения вычислительных ресурсов многопроцессорных систем (МС), базирующихся на принципах распараллеливания и конвейеризации, является одной из центральных при создании эффективного системного и прикладного программного обеспечения [1]. Важное место в ее решении отводится оптимальной организации конкурирующих процессов, использующих общие программные ресурсы, поскольку от этого зависит не только эффективность использования МС, но и предоставляемые ими возможности решения в реальное время сложных задач из различных областей знаний. В связи с этим всевозрастающую роль приобретают МС на базе распределенной обработки, поэтому особую актуальность имеют задачи построения и исследования математических моделей оптимальной организации конкурирующих процессов при распределенной обработке.

1. Математическая модель конкурирующих процессов при распределенной обработке

Как и в [2], математическая модель распределенной обработки конкурирующих процессов включает в себя $p \geq 2$ процессоров многопроцессорной системы, $n \geq 2$ конкурирующих процессов, $s \geq 2$ блоков Q_j, $j = \overline{1,s}$, структурированного на блоки программного ресурса, матрицу $[t_{ij}]$, $i = \overline{1,n}$, $j = \overline{1,s}$, времени выполнения блоков конкурирующими процессами, параметр $\varepsilon > 0$, характеризующий время (накладные расходы), затрачиваемое на организацию параллельного использования блоков программного ресурса множеством распределенных конкурирующих процессов. Предполагается, что все n процессов используют одну копию структурированного на блоки программного ресурса, причем из физических соображений на множестве блоков установлен линейный порядок их выполнения.

В [2] введены базовые режимы взаимодействия процессов, процессоров и блоков структурированного программного ресурса:

- *асинхронный режим*, который предполагает отсутствие простоев процессоров при условии готовности блоков, а также невыполнение блоков при наличии процессоров;
- *первый синхронный режим*, обеспечивающий непрерывное выполнение блоков программного ресурса внутри каждого из процессов;
- *второй синхронный режим*, обеспечивающий непрерывное выполнение каждого блока всеми процессами.

Определение 1. Система конкурирующих процессов называется *одинаково распределенной*, если с учетом накладных расходов времена t^{ε}_{ij} выполнения блоков Q_j, $j = \overline{1,s}$, программного ресурса каждым из i-х процессов совпадают и равны $t_i^\varepsilon = t_i + \varepsilon$ для всех $i = \overline{1,n}$, т. е. справедлива цепочка равенств $t_{i1}^\varepsilon = t_{i2}^\varepsilon = ... = t_{is}^\varepsilon = t_i^\varepsilon$, $i = \overline{1,n}$.

2. Оптимальность одинаково распределенных систем

В [2] доказано, что для одинаково распределенных конкурирующих процессов с учетом накладных расходов $\varepsilon > 0$ при достаточном числе процессоров, т. е. $2 \le s \le p$, для всех трех базовых режимов минимальное общее время выполнения $T(p,n,s,\varepsilon)$ совпадает и вычисляется по формуле

$$T(p,n,s,\varepsilon) = T_\varepsilon^n + (s-1)t_{\max}^\varepsilon, \tag{1}$$

где $T_\varepsilon^n = \sum_{i=1}^{n} t_i^\varepsilon$, $t_{\max}^\varepsilon = \max_{1 \le i \le n} t_i^\varepsilon$, $t_i^\varepsilon = t_i + \varepsilon$.

Выделим в классе одинаково распределенных систем конкурирующих процессов специальный подкласс так называемых *стационарных* систем.

Определение 2. Одинаково распределенную систему конкурирующих процессов назовем *стационарной*, если $t_1^\varepsilon = t_2^\varepsilon = ... = t_n^\varepsilon = t^\varepsilon$.

Используя (1), нетрудно показать, что для всех трех базовых режимов в случае стационарной одинаково распределенной системы конкурирующих процессов при достаточном числе процессоров минимальное общее время их выполнения $\overline{T}_\varepsilon$ определяется равенством

$$\overline{T}_\varepsilon(s \le p) = (n+s-1)t_\varepsilon, \tag{2}$$

где $t_\varepsilon = T^n / n + \varepsilon$, $T^n = nt$.

Определение 3. Одинаково распределенную систему конкурирующих процессов будем называть *эффективной* при фиксированных p, $s \ge 2$, если выполняется соотношение $\Delta_\varepsilon = sT^n - T(p,n,s,\varepsilon) \ge 0$, 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 Δ_ε первой системы не меньше соответствующей величины второй. Для введенного подмножества одинаково распределенных систем справедливо следующее утверждение [2].

Теорема 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.*

Для любой стационарной одинаково распределенной системы с учетом (2) имеем

$$\overline{\Delta}_\varepsilon(s \le p) = (s-1)(T^n - t) - (n+s-1)\varepsilon \ge 0, \tag{3}$$

где $t = T^n / n$.

Определение 4. Эффективная одинаково распределенная система называется *оптимальной*, если величина Δ_ε достигает наибольшего значения.

В [3] показано, что оптимальную одинаково распределенную систему достаточно искать среди эффективных одинаково распределенных систем. Более того, в силу теоремы 1 оптимальную одинаково распределенную систему следует искать среди стационарных одинаково распределенных систем. Тогда с учетом (3) имеем

$$\overline{\Delta}_\varepsilon(s \le p) = (s-1)T^n(1 - 1/n) - (n+s-1)\varepsilon.$$

Введем функцию действительного аргумента x вида

$$\overline{\Delta}_\varepsilon(x) = (s-1)T^n\left(1 - \frac{1}{x}\right) - (x+s-1)\varepsilon, \ x \ge 1.$$

Решение задачи об оптимальности одинаково распределенной системы, состоящей из n конкурирующих процессов, для достаточного числа процессоров для всех трех базовых режимов следует из теоремы.

Теорема 2. *Для того чтобы эффективная одинаково распределенная система конкурирующих процессов была оптимальной при заданных* $2 \le s \le p$, 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.*

Доказательство. *Необходимость.* Рассмотрим введенную функцию

$$\overline{\Delta}_\varepsilon(x)=(s-1)T^n\left(1-\frac{1}{x}\right)-(x+s-1)\varepsilon,\ x\ge 1.$$

Согласно определению 4 одинаково распределенная система будет оптимальной в той точке x, где функция $\overline{\Delta}_\varepsilon(x)$ достигает своего наибольшего значения. Покажем, что функция $\overline{\Delta}_\varepsilon(x)$ достигает своего наибольшего значения в точке $x=\sqrt{\frac{(s-1)T^n}{\varepsilon}}$. Действительно,

$$\overline{\Delta}'_\varepsilon(x)=\frac{(s-1)T^n}{x^2}-\varepsilon,\ \ \overline{\Delta}''_\varepsilon(x)=-\frac{2T^n(s-1)}{x^3}<0,\ \text{так как}\ s\ge 2,\ x>0.$$

Следовательно, функция $\overline{\Delta}_\varepsilon(x)$ достигает максимума в точке, где первая ее производная обращается в нуль $\overline{\Delta}'_\varepsilon(x)=0$, т. е. $x^*=\sqrt{\frac{(s-1)T^n}{\varepsilon}}$.

Целочисленными точками, в которых достигается наибольшее значение функции $\overline{\Delta}_\varepsilon(x)$, будут $n_0=[x^*]$ или $n_0=[x^*]+1$. Следовательно, в качестве n_0 можно выбрать одно из чисел $\left[\sqrt{\frac{(s-1)T^n}{\varepsilon}}\right]$, $\left[\sqrt{\frac{(s-1)T^n}{\varepsilon}}\right]+1$, в которых функция $\overline{\Delta}_\varepsilon(s\le p)$ принимает наибольшее значение.

Если же окажется, что ни одна из точек $\left[\sqrt{\frac{(s-1)T^n}{\varepsilon}}\right]$, $\left[\sqrt{\frac{(s-1)T^n}{\varepsilon}}\right]+1$, в которой функция $\overline{\Delta}_\varepsilon(x)$ принимает наибольшее значение, не принадлежит $[2,n]$, то в качестве оптимальной выбираем эффективную одинаково распределенную систему с числом процессов $n_0=n$.

В силу отрицательности второй производной исследуемая функция выпукла. Следовательно, точка максимума всегда существует, а значит, существует и эффективная одинаково распределенная система конкурирующих процессов в случае, когда $n\to\infty$.

Достаточность следует из свойств выпуклости функции $\overline{\Delta}_\varepsilon(x)$ при $s\le p$ на отрезке $[2,n]$.

В случае ограниченного параллелизма с учетом параметра $\varepsilon>0$, характеризующего время дополнительных системных расходов на организацию параллельного использования блоков множеством распределенных конкурирующих процессов, для вычисления минимального общего времени в асинхронном и втором синхронном режимах для класса одинаково распределенных конкурирующих процессов имеют место формулы [2]:

$$T(p,n,s,\varepsilon)=\begin{cases}kT_\varepsilon^n+(p-1)t_{\max}^\varepsilon\ \textit{при}\ s=kp,\ k>1,\ T_\varepsilon^n>pt_{\max}^\varepsilon,\\(k+1)T_\varepsilon^n+(r-1)t_{\max}^\varepsilon\ \textit{при}\ s=kp+r,\ k\ge 1,\ 1\le r<p,\ T_\varepsilon^n>pt_{\max}^\varepsilon.\end{cases}\tag{4}$$

Для решения задачи об оптимальности одинаково распределенной системы конкурирующих процессов в случае ограниченного параллелизма в асинхронном и втором синхронном режимах с учетом (4) введем функции действительного аргумента x вида

$$\overline{\Delta}_\varepsilon(x)=(s-1)T^n-\frac{(p-1)T^n}{x}-(kx+p-1)\varepsilon\ \text{при}\ s=kp,\ k>1,\tag{5}$$

$$\overline{\Delta}_\varepsilon(x)=(s-k-1)T^n-\frac{(r-1)T^n}{x}-\big((k+1)x+(r-1)\big)\varepsilon\tag{6}$$

при $s=kp+r,\ k\ge 1,\ 1\le r<p.$

Теорема 3. *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_0 *в системе равнялось одному из чисел:*

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

2) $\left[\left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right], \left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right]+1\right] \cap [2,n]$ 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*,* n *– заданное число.*

Доказательство. *Необходимость.* Для случая $s = kp,\ k > 1,$ функция вида (5) достигает наибольшего значения в точке $x^* = \sqrt{\dfrac{(p-1)T^n}{k\varepsilon}}$, так как

$$\overline{\Delta}'_\varepsilon(x) = \frac{(p-1)T^n}{x^2} - k\varepsilon.$$

Как и в случае неограниченного параллелизма, целочисленными точками будут $n_0 = [x^*]$ или $n_0 = [x^*] + 1$. Следовательно, в качестве оптимальной выбираем эффективную одинаково распределенную систему с числом процессов $\left[\sqrt{\dfrac{(p-1)T^n}{k\varepsilon}}\right], \left[\sqrt{\dfrac{(p-1)T^n}{k\varepsilon}}\right]+1.$

При $s = kp + r,\ k \ge 1,\ 1 \le r < p,$ первая производная функции (6) имеет вид

$$\overline{\Delta}'_\varepsilon(x) = \frac{(r-1)T^n}{x^2} - (k+1)\varepsilon.$$

Следовательно, в качестве n_0 можно выбрать одно из значений $\left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right]$ или $\left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right]+1.$

Если же окажется, что ни одна из точек $\left[\sqrt{\dfrac{(p-1)T^n}{k\varepsilon}}\right],$ $\left[\sqrt{\dfrac{(p-1)T^n}{k\varepsilon}}\right]+1,$ $\left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right],$ $\left[\sqrt{\dfrac{(r-1)T^n}{(k+1)\varepsilon}}\right]+1,$ в которых функции (5), (6) принимают наибольшее значение, не принадлежит $[2,n]$, то в случае ограниченного параллелизма в качестве оптимальной выбираем эффективную одинаково распределенную систему с числом процессов $n_0 = n.$

Достаточность следует из свойств выпуклости функции $\overline{\Delta}_\varepsilon(x)$ при $s > p$ на отрезке $[2,n]$. Действительно, исследуемые функции (5) и (6) выпуклы в силу отрицательности вторых производных:

$$1)\ \overline{\Delta}''_\varepsilon(x) = -\frac{2T^n(p-1)}{x^3} < 0,\quad p \ge 2,\ x > 0,$$

$$2)\ \overline{\Delta}''_\varepsilon(x) = -\frac{2T^n(r-1)}{x^3} < 0,\ 1 \le r < p,\ x > 0.$$

Теорема доказана.

* * *

Полученные критерии оптимальности систем одинаково распределенных конкурирующих процессов могут быть использованы при проектировании системного и прикладного программного обеспечения для многопроцессорных систем и сетей при решении проблем оптимального использования вычислительных ресурсов.

1. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. СПб., 2002.
2. Коваленко Н.С., Павлов П.А. // Дакл. НАН Беларусі. Сер. фіз.-мат. навук. 2005. № 6. С. 32.
3. Павлов П.А. // Вестн. Фонда фундамент. исслед. 2006. № 1. С. 55.

Поступила в редакцию 10.12.08.

Павел Александрович Павлов – кандидат физико-математических наук, доцент кафедры высшей математики и информационных технологий Полесского государственного университета.