УДК 681.324

Г.А. Кучук, А.А. Пашнев, Л.Н. Бондарь

ПОИСК РАЦИОНАЛЬНОГО РАСПРЕДЕЛЕНИЯ ЗАДАЧ ПО УЗЛАМ ВЫЧИСЛИТЕЛЬНОЙ СЕТИ В ПРОЦЕССЕ ИХ РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ

Постановка проблемы в общем виде. Используя постановку задачи, представленную в [1 - 3] сформулируем проблему поиска рационального распределения задач по узлам вычислительной сети в процессе их распределенной обработки в следующем виде: необходимо найти такое разбиение множества задач, решаемых в вычислительной сети (ВС) на подмножества и их распределение по узлам ВС в процессе распределенной обработки, при котором средняя задержка пакета данных в сети принимала минимальное значение.

Целью статьи является разработка алгоритма поиска рационального распределения задач по узлам вычислительной сети, обеспечивающего рациональное разбиение множества задач, решаемых в ВС на подмножества и их распределение по узлам сети в процессе распределенной обработки, позволяющее минимизировать среднюю задержку пакета данных в ней.

Целевая функция задачи поиска рационального разбиения множества задач Z, обрабатываемых в вычислительной сети на подмножества и их распределения по узлам $y_a \in Y$, полученная в [1] определяется выражением

$$F^{(\gamma)} = \frac{1}{u_{z_{\max}}} \cdot \sum_{b=1}^{h_z} \sum_{a=1}^{h_y} m_{z_{b,a}} \cdot s_{z_{b,a}}, \qquad (1)$$

где $u_{z_{\max}}$ - независимая от распределения γ величина, определяющая максимальную суммарную интенсивность обмена задач с узлами вычислительной сети в соответствии с выражением $u_{z_{\max}} = \sum_{b=1}^{h_z} \sum_{i=1}^{h_y} u_{z_{b,i}}$; h_z - число обрабатываемых в ВС задач; h_y - число узлов вычислительной сети; $u_{z_{b,i}}$ - интенсивность обмена задачи $z_b \in Z$ с узлом $y_i \in Y$; $m_{z_{b,a}}$ - требуемый вычислительный ресурс узла $y_a \in Y$, необходимый для выполнения подзадачи задачи $z_b \in Z$; $s_{z_{b,a}}$ -штраф при распределении подзадачи задачи $z_b \in Z$ на узел $y_a \in Y$, определяемый выражением $s_{z_{b,a}} = \sum_{i=1}^{h_y} \left(u_{z_{b,i}} \cdot h_{wa,i} \right) / \varphi_{z_b}$; $h_{wa,i}$ - длина кратчайшего маршрута между узлами y_a и y_i, определяемая числом каналов передачи данных, входящих в этот маршрут; φ_{z_b} - требуемый вычислительный ресурс для обработки задачи $z_b \in Z$.

Искомое распределение γ должно удовлетворять следующим условиям [3]:

1) $\forall y_a \in Y \left| \sum_{b=1}^{h_z} m_{z_{b,a}} \le \varphi_{y_a} \right.$;

2) $\forall z_b \in Z \left| \sum_{a=1}^{h_y} m_{z_{b,a}} \le \varphi_{z_b} \right.$;

3) $\sum_{a=1}^{h_y} \varphi_{y_a} \ge \sum_{b=1}^{h_z} \varphi_{z_b}$;

4) $\mathrm{s}_{z_{b,a}} \ge 0, \mathrm{m}_{z_{b,a}} \ge 0$ для $1 \le a \le h_y$, $1 \le b \le h_z$,

где φ_{y_a} - доступный вычислительный ресурс узла $y_a \in Y$ (производительность его ПЭВМ).

С учетом приведенных условий, задача поиска рационального разбиения множества задач Z, обрабатываемых в вычислительной сети на подмножества и их распределения по узлам $\mathrm{y_a} \in \mathrm{Y}$ может быть сформулирована следующим образом.

Пусть заданы множества задач Z и узлов Y вычислительной сети, определяемые кортежами

$$\langle Z, \varphi_z, U_z \rangle \text{ и } \langle Y, \varphi_y, H_w \rangle,$$

где $\varphi_z = (\varphi_{z_1}, ..., \varphi_{z_{h_z}})$ - вектор требуемых вычислительных ресурсов для обработки множества задач Z; $U_z = \left\| u_{z_{b,i}} \right\|$ - матрица интенсивностей обмена задач множества Z с узлами множества Y; $\varphi_y = (\varphi_{y_1}, ..., \varphi_{y_{h_y}})$ - вектор доступных вычислительных ресурсов множества узлов Y вычислительной сети; $H_w = \left\| h_{w_{a,i}} \right\|$ - матрица длин кратчайших маршрутов между каждой парой узлов ВС y_a и y_i, $1 \le a \le h_y$, $1 \le i \le h_y$. Требуется найти такое распределение γ, удовлетворяющее условиям 1 - 4, чтобы выражение (1) принимало минимальное значение.

При построении алгоритма решения сформулированной задачи удобно принять, что общий суммарный доступный вычислительный ресурс узлов множества Y равен общему суммарному требуемому вычислительному ресурсу задач множества Z, т.е. $\sum_{a=1}^{h_y} \varphi_{y_a} = \sum_{b=1}^{h_z} \varphi_{z_b}$.

С этой целью необходимо ввести фиктивный (h_y+1)-й узел с доступным вычислительным ресурсом $\varphi_{y_{h_y+1}}$ либо фиктивную (h_z+1)-ю задачу с требуемым вычислительным ресурсом $\varphi_{z_{h_z+1}}$ при которых

$$\sum_{a=1}^{h_y+1} \varphi_{y_a} = \sum_{b=1}^{h_z+1} \varphi_{z_b} ,$$

и принять штраф:

$$\mathrm{s}_{z_{h_z+1,a}} = 0,\ 1 \le a \le h_y,\ \text{и}\ \mathrm{s}_{z_{b,h_y+1}} = \max_{\substack{1\le b\le h_z \\ 1\le a\le h_y}} s_{z_{b,a}} ,\ 1 \le b \le h_z .$$

Рассмотрим пошагово алгоритм, осуществляющий поиск рационального распределения задач по узлам вычислительной сети, обеспечивающий разбиение множества задач, решаемых в ВС на подмножества и их распределение по узлам сети, позволяющее минимизировать среднюю задержку пакета данных в ней в процессе распределенной обработки задач.

Шаг 1.

Базовое разбиение множества задач, решаемых в ВС на подмножества и их распределение по узлам сети, определяется матрицей $M_z^{(\gamma)}$, заполненной по правилу минимальных штрафов [4], элемент $\mathrm{m}_{z_{b,a}}$ которой определяет вычислительный ресурс узла $y_a \in Y$, выделенный для обработки подзадачи задачи $z_b \in Z$. Для построения базового распределения формируется матрица штрафов S_z, элемент $\mathrm{s}_{z_{b,a}}$ которой определяет штраф при выделении подзадаче задачи z_b узлом y_a единицы вычислительного ресурса.

Суть правила минимальных штрафов заключается в том, что заполнение матрицы $M_z^{(\gamma)}$ начинается последовательно с элемента $\mathrm{m}_{z_{b,a}}$, которому в матрице штрафов S_z соответствует наименьшее значение штрафа $\mathrm{s}_{z_{b,a}}$. Элементу $\mathrm{m}_{z_{b,a}}$ присваивается значение $\min(\varphi_{z_b}, \varphi_{y_a})$. Изменяются значения требуемого вычислительного ресурса для обработки задачи z_b и доступного вычислительного ресурса узла y_a:

$$\varphi_{z_b} = \varphi_{z_b} - \min(\varphi_{z_b}, \varphi_{y_a});$$

$$\varphi_{y_a} = \varphi_{y_a} - \min(\varphi_{z_b}, \varphi_{y_a}).$$

Из дальнейшего рассмотрения в матрице S_z исключается либо строка, соответствующая заявке z_b, потребность в вычислительных ресурсах которой полностью удовлетворена, либо столбец, соответствующий узлу y_a доступный вычислительный ресурс которого полностью израсходован, либо столбец и строка, если полностью израсходован доступный вычислительный ресурс узла y_a и удовлетворена потребность в вычислительных ресурсах заявки z_b.

Из оставшихся элементов матрицы S_z снова последовательно выбираются элементы с наименьшим значением штрафа, и процесс распределения вычислительных ресурсов продолжается до тех пор, пока потребности всех заявок множества Z не будут удовлетворены в вычислительных ресурсах. Незаполненным элементам матрицы $M_z^{(\gamma)}$ присваиваются нулевые значения.

Шаг 2.

Для проверки базового распределения на оптимальность строится система потенциалов. Систему потенциалов можно построить только для невырожденного плана распределения. План распределения γ является невырожденным, если число элементов матрицы $M_z^{(\gamma)}$, отличных от нуля, равно $h_z + h_y - 1$. [5] При построении базового распределения может оказаться, что не нулевых элементов матрицы $M_z^{(\gamma)}$ меньше, чем $h_z + h_y - 1$, т.е. план распределения γ является вырожденным. В этом случае нулевым элементам матрицы $M_z^{(\gamma)}$ с наименьшим значением $s_{z_{b,a}}$ последовательно присваиваются значения условного базисного нуля до тех пор, пока число элементов матрицы $M_z^{(\gamma)}$, отличных от нуля не станет равно $h_z + h_y - 1$.

Каждому узлу y_a ставится в соответствие потенциал p_{y_a}, а каждой заявке z_b – потенциал p_{z_b}. Потенциалы p_{y_a} и p_{z_b} выбираются, так, чтобы для каждого не нулевого элемента матрицы $M_z^{(\gamma)}$ выполнялось условие $p_{z_b} + p_{y_a} = s_{z_{b,a}}$, а для нулевого элемента матрицы $M_z^{(\gamma)}$ - условие $p_{z_b} + p_{y_a} \le s_{z_{b,a}}$. Так как число всех потенциалов равно $h_z + h_y$, а отличных от нуля элементов матрицы $M_z^{(\gamma)}$ $h_z + h_y - 1$, то для определения значений p_{z_b} и p_{y_a} решается система уравнений $p_{z_b} + p_{y_a} = s_{z_{b,a}}$ с $h_z + h_y$ неизвестными, для чего одному из неизвестных присваивается произвольное значение, и система имеет единственное решение.

Шаг 3.

Для каждого из нулевых элементов матрицы $M_z^{(\gamma)}$ вычисляется значение его базисной оценки, определяемой разностью между величиной штрафа $\mathrm{s}_{z_{b,a}}$ и суммой потенциалов $p_{z_b} + p_{y_a}$, соответствующих данному элементу. План является рациональным в том случае, если полученные значения базисных оценок $\mathrm{r}_{s_{b,a}}$ для всех нулевых элементов $\mathrm{m}_{z_{b,a}}$ матрицы $M_z^{(\gamma)}$ будут не отрицательными, т.е. $\mathrm{r}_{s_{b,a}} = \mathrm{s}_{z_{b,a}} - \left(p_{z_b} + p_{y_a}\right) \geq 0$.

Если хотя бы одна из вычисленных базисных оценок имеет отрицательное значение, осуществляется текущее перераспределение вычислительных ресурсов. Для этого, из матрицы $M_z^{(\gamma)}$ выбирается элемент $m_{z_{b,a}}$, для которого базисная оценка $\mathrm{r}_{s_{b,a}}$ принимает минимальное отрицательное значение. Для выбранного элемента $\mathrm{m}_{z_{b,a}}$ в матрице $M_z^{(\gamma)}$ строится замкнутый контур. Замкнутый контур представляет собой последовательность элементов матрицы $M_z^{(\gamma)}$, два соседних элемента которой расположены в одной строке или в одном столбце, и последний элемент находится в той же строке или столбце, что и первый.

Шаг 4.

Нулевой элемент $m_{z_{b,a}}$ матрицы $M_z^{(\gamma)}$, для которого $\mathrm{r}_{s_{b,a}}$ принимает минимальное отрицательное значение, выбирается первым элементом замкнутого контура. Остальные элементы замкнутого контура выбираются из элементов матрицы $M_z^{(\gamma)}$ отличных от нуля.

Выбрав в качестве исходной одну из координат первого элемента контура, например, координату столбца матрицы $M_z^{(\gamma)}$, анализируются строки матрицы $M_z^{(\gamma)}$ в этом столбце. Следующий, отличный от нуля элемент матрицы $M_z^{(\gamma)}$ в рассматриваемом столбце выбирается в качестве второго элемента контура. Третий элемент контура ищется в строке, в которой находится второй выбранный элемент. Если в этой строке отсутствует не нулевой элемент, то второй найденный элемент удаляется из контура, и в столбце матрицы $M_z^{(\gamma)}$, заданном первым элементом контура, ищется другой ненулевой элемент, который и будет вторым элементом контура. Координата строки найденного элемента указывает строку матрицы $M_z^{(\gamma)}$, в которой ищется следующий ненулевой элемент.

Процесс поиска элементов замкнутого контура продолжается до тех пор, пока не совпадут координаты строки последнего найденного элемента и первого элемента контура.

Элементам построенного замкнутого контура приписываются чередующиеся знаки, причем первому элементу контура приписывается положительный знак. Среди элементов контура с отрицательным знаком выбирается элемент $m_{z_{b,a}}$, которому соответствует наименьшее значение. Значение элемента $m_{z_{b,a}}$ прибавляется к значениям элементов замкнутого контура с положительным знаком и вычитается из значений элементов замкнутого контура с отрицательным знаком. В результате получим новое разбиение множества задач, решаемых в вычислительной сети на подмножества и их распределение по узлам ВС.

Для нового распределения γ вновь строится система потенциалов, и вычисляются значения базисных оценок для нулевых элементов матрицы $M_z^{(\gamma)}$. В случае отсутствия отрицательных значений базисных оценок для нулевых элементов матрицы $M_z^{(\gamma)}$ в распределении γ, принимается решение, о том, что, полученное разбиение множества задач, решаемых в ВС на подмножества и их распределение по узлам сети рационально.

Выводы. Таким образом, основным, полученным научным и практическим результатом данного исследования является разработанный алгоритм поиска рационального распределения задач по узлам вычислительной сети, обеспечивающий рациональное разбиение множества задач, решаемых в ВС на подмножества и их распределение по узлам сети, позволяющее минимизировать среднюю задержку пакета данных в ней в процессе распределенной обработки задач.

1. *Пашнев А.А., Клименко Л.А.* Математическая модель задачи рационального управления распределенной обработкой задач в информационно-телекоммуникационной сети // Системи обробки інформації. – Х.: ХВУ. – 2004. – Вип. 3. – С. 162 – 168.
2. *Королев А.В., Кучук Г.А., Пашнев А.А.* Адаптивная маршрутизация в корпоративных сетях. – Х.: ХВУ, 2003. – 228с.
3. *Королев А.В., Кучук Г.А., Пашнев А.А.* Управление сетевыми ресурсами. – Х.: ХВУ, 2004. – 272с.
4. *Майника Э.* Алгоритмы оптимизации на сетях и графах: Пер. с англ. / Под ред. Е.К. Масловского. – М.: Мир, 1981. – 321с.
5. *Таха Х.* Введение в исследование операций. – М.: Мир, 1985. - Ч.1. – 479с.

Поступила 16.06.2004 г.