УДК 681.324

РАСПРЕДЕЛЕНИЕ ВЫЧИСЛИТЕЛЬНОГО РЕСУРСА ОДНОРОДНОЙ ВЫЧИСЛИТЕЛЬНОЙ СЕТИ ПО КВАНТАМ ЗАДАННОГО ИНТЕРВАЛА ВРЕМЕНИ

к.т.н. А.А. Пашнев, к.т.н. Г.А. Кучук, И.А. Лебедева
(представил д.т.н., проф. А.В. Королёв)

Предлагается математическая модель, позволяющая построить равномерное распределение выделенного вычислительного ресурса для обработки множества заявок абонентов однородной вычислительной сети по квантам заданного интервала времени.

Постановка задачи в общем виде. Для обеспечения максимальной загрузки однородной вычислительной сети (ОВС) и удовлетворения обрабатываемых в ней заявок на требуемый вычислительный ресурс в заданный для каждой заявки интервал времени, возникает задача составления такого допустимого расписания выполнения заданного комплекса задач по заявкам абонентов, которое обеспечит возможность применения ПЭВМ с минимально необходимой производительностью [1]. Подобные задачи рассматривались в [2, 3], но с другой целевой функцией и без учета равномерности распределения выделенного вычислительного ресурса.

В связи с этим, **целью статьи** является разработка математической модели, позволяющей построить равномерное распределение выделенного вычислительного ресурса для обработки множества заявок абонентов ОВС по квантам заданного интервала времени.

Исходными данными для решения задачи распределения выделенного вычислительного ресурса по квантам заданного временного интервала при обработке заявок абонентов в однородной вычислительной сети являются: T_z – заданный интервал времени; Z – множество заявок абонентов ОВС. Пусть интервал времени T_z представляет собой конечный набор квантов, равных 1 с. Выбирая естественную нумерацию этих квантов, представим временной интервал T_z в виде отрезка натурального ряда $\left\{ t_{z_1}, t_{z_2}, \ldots, t_{z_i}, \ldots, t_{z_{h_t}} \right\}$, где t_{z_i} обозначает i-й квант времени, $1 \le i \le h_t$; h_t – число квантов временного интервала T_z. Каждая заявка $z_b \in Z$, $1 \le b \le h_z$, характеризуется параметрами φ_{z_b}, T_{z_b}, где h_z – число заявок абонентов, обрабатываемых в ОВС; φ_{z_b} – требуемый вычислительный ресурс для обработки заявки z_b; $T_{z_b} = \left\{ t_{z_{b_1}}, t_{z_{b_2}} \right\}$ – интервал времени, в течении которого необходимо предоставить требуемый вычислительный ресурс; $t_{z_{b_1}}$ – начальный квант временного интервала T_{z_b}; $t_{z_{b_2}}$ – конечный квант временного интервала T_{z_b}. Считая, что каждая заявка соответствует одной задаче, получим $\varphi_{z_b} = \omega_{z_b}$, где ω_{z_b} – необходимый объем вычислений (число машинных операций) для выполнения задачи z_b.

В результате распределения γ вычислительного ресурса сети формируется матрица $M_{\varphi}^{(\gamma)}$, в которой каждой заявке $z_b \in Z$ сопоставляется вектор-строка $m_{\varphi_b} = (m_{\varphi_{b,1}}, \ldots, m_{\varphi_{b,h_t}})$, представляющая собой расписание выделения вычислительного ресурса ОВС для обработки заявки z_b, где компонент $m_{\varphi_{b,i}}$ определяет выделенный для заявки z_b вычислительный ресурс в i-й квант времени.

Качество распределения γ можно оценить с помощью целевой функции $F^{(\gamma)}$ и величины максимального суммарного выделенного вычислительного ресурса, приходящегося на квант заданного временного интервала T_Z в распределении γ по всем заявкам множества Z:

$$m_{\varphi_{\max}}^{(\gamma)} = \max_{i=1,\ldots,h_t} \sum_{b=1}^{h_z} m_{\varphi_{b,i}} .$$

При условии равномерного распределения по квантам заданного временного интервала T_Z суммарного объема вычислений, необходимого для обработки заявок множества Z, выражение для определения величины минимального суммарного требуемого вычислительного ресурса, приходящегося на квант заданного временного интервала T_Z, примет следующий вид:

$$\varphi_{z_{\min}} = \frac{1}{h_t} \sum_{b=1}^{h_z} \varphi_{z_b} .$$

Основой определения целевой функции $F^{(\gamma)}$ служит штраф при выделении заявке $z_b \in Z$ единицы вычислительного ресурса в i-й квант времени. Если единица вычислительного ресурса для заявки z_b, характеризующейся интервалом времени обработки $T_{z_b} = \left\{ t_{z_{b_1}}, t_{z_{b_2}} \right\}$, выделена в i-й квант времени, то соответствующий ей штраф определяется с помощью выражения

$$s_{z_{b,i}} = \begin{cases} 0, & \text{если} \quad t_{z_{b_1}} \le t_{z_i} \le t_{z_{b_2}}; \\ \left(t_{z_{b_1}} - t_{z_i}\right) / \varphi_{z_b}, & \text{если} \quad t_{z_i} < t_{z_{b_1}}; \\ \left(t_{z_i} - t_{z_{b_2}}\right) / \varphi_{z_b}, & \text{если} \quad t_{z_i} > t_{z_{b_2}}. \end{cases}$$

Таким образом, для каждой заявки $z_b \in Z$ имеем вектор $\mathbf{s}_{z_b} = \left(\mathbf{s}_{z_{b_1}}, \ldots, \mathbf{s}_{z_{b_{h_t}}}\right)$, у которого компонент $\mathbf{s}_{z_{b_i}}$, $1 \le i \le h_t$, определяет величину штрафа при выделении заявке z_b единицы вычислительного ресурса в i-й квант времени.

Величина штрафа, характеризующего полученное распределение γ выделенного вычислительного ресурса для обработки множества заявок Z, определяет целевую функцию $F^{(\gamma)} = \sum_{b=1}^{h} \sum_{i=1}^{h_t} m_{\varphi_{b,i}} \cdot s_{z_{b,i}}$.

При построении распределения γ вычислительного ресурса по квантам на заданном интервале времени T_Z необходимо минимизировать величины $F^{(\gamma)}$ и $m^{(\gamma)}_{\varphi_{max}}$.

Дадим математическое описание задачи построения равномерного распределения выделенного вычислительного ресурса для обработки множества заявок абонентов по квантам заданного интервала времени.

Пусть $\langle Z, \varphi_Z, T_Z \rangle$ – кортеж заявок абонентов ОВС, где Z – множество заявок; $\varphi_Z : Z \to N_+$ – функция, указывающая каждой заявке $z_b \in Z$ требуемый вычислительный ресурс для ее обработки; $T_Z : Z \to N_+$ – функция, указывающая каждой заявке $z_b \in Z$ интервал времени для ее обработки, где $T_{z_b} = \left\{ t_{z_{b_1}}, t_{z_{b_2}} \right\}$ – пара, указывающая заявке z_b начальный $t_{z_{b_1}}$ и конечный $t_{z_{b_2}}$ кванты времени ее обработки, при этом $t_{z_{b_1}} < t_{z_{b_2}}$.

Полученное распределение γ можно описать с помощью кортежа

$$\left\langle Z, \varphi_Z, T_Z, \varphi_y, M^{(\gamma)}_{\varphi}, F^{(\gamma)}, m^{(\gamma)}_{\varphi_{max}} \right\rangle,$$

где $\varphi_y : T_Z \to N_+$ – функция, указывающая каждому кванту времени $t_{z_i} \in T_Z$ доступный суммарный вычислительный ресурс ОВС. При этом распределение γ должно удовлетворять следующим условиям:

1) $\forall z_b \in Z, \forall t_{z_i} \in T_{z_b}$ $\quad m_{\varphi_{b,i}} \geq 0, s_{z_{b,i}} \geq 0$;

2) $\forall z_b \in Z$ $\quad \sum_{i=1}^{h_t} m_{\varphi_{b,i}} \leq \varphi_{z_b}$;

3) $\forall t_{z_i} \in T_z$ $\quad \sum_{b=1}^{h_z} m_{\varphi_{b,i}} \leq \varphi_{y_i}$.

Таким образом, задача построения равномерного распределения вычислительного ресурса по квантам заданного интервала времени обработки множества заявок с учетом приведенных выше понятий и обозначений может быть сформулирована следующим образом.

Пусть задано множество заявок абонентов ОВС, характеризующееся кортежем $\langle Z, \varphi_Z, T_Z \rangle$. Требуется найти такое распределение вычислительного ресурса, чтобы величины $F^{(\gamma)}$ и $m_{\varphi_{max}}^{(\gamma)}$ принимали минимальное значение при выполнении условий 1 – 3. Поставим в соответствие каждому i-му кванту заданного интервала времени T_z доступный суммарный вычислительный ресурс ОВС φ_{y_i}, а каждой заявке $z_b \in Z$ – требуемый вычислительный ресурс φ_{z_b} для ее обработки. Штраф при выделении единицы вычислительного ресурса в i-й квант времени для обработки заявки z_b – компонент $\mathbf{s}_{z_{b_i}}$ вектора $\mathbf{s}_{z_b}$. Выделенный вычислительный ресурс для обработки заявки z_b в i-й квант времени определяется величиной $m_{\varphi_{b,i}}$.

Для получения допустимого решения задачи распределения вычислительного ресурса по квантам заданного временного интервала необходимо обеспечить выполнение неравенства $\sum_{i=1}^{h_t} \varphi_{y_i} \geq \sum_{b=1}^{h_z} \varphi_{z_b}$. При построении алгоритма решения рассматриваемой задачи удобно принять, что доступный суммарный вычислительный ресурс ОВС по всем квантам заданного интервала времени T_Z должен быть равен суммарному требуемому вычислительному ресурсу по всем заявкам множества Z, т.е. $\sum_{i=1}^{h_t} \varphi_{y_i} = \sum_{b=1}^{h_z} \varphi_{z_b}$. С этой целью достаточно ввести фиктивный $h_t + 1$-й квант времени с доступным суммарным вычислительным ресурсом ОВС, равным $\varphi_{y_{h_t+1}} = \sum_{b=1}^{h_z} \varphi_{z_b} - \sum_{i=1}^{h_t} \varphi_{y_i}$, либо фиктивную заявку z_{h_z+1} с требуемым вычислительным ресурсом, равным $\varphi_{z_{h_z+1}} = \sum_{i=1}^{h_t} \varphi_{y_i} - \sum_{b=1}^{h_z} \varphi_{z_b}$, при которых $\sum_{i=1}^{h_t+1} \varphi_{y_i} = \sum_{b=1}^{h_z+1} \varphi_{z_b}$. При этом штраф $s_{z_{h_z+1,i}} = 0$, $1 \le i \le h_t$. При вычислении величины штрафа $s_{z_{b,h_t+1}}$, $1 \le b \le h_z$, воспользуемся следующим правилом. В случае, если для любой заявки $z_b \in Z$ вычислительный ресурс должен быть выделен только в соответствующем временном интервале T_{z_b}, то $s_{z_{b,h_t+1}} = \left(t_{z_{h_t+1}} - t_{z_{b_2}}\right) / \varphi_{z_b}$. Если же для любой заявки $z_b \in Z$ допускается выделение вычислительного ресурса на всех квантах заданного временного интервала T_z, то $s_{z_{b,h_t+1}} = T_z / \varphi_{z_b}$.

Рассмотренная задача имеет $h_z + h_t + 1$ переменных. Алгоритм решения данной задачи включает:

1) построение базового распределения вычислительного ресурса;
2) построение системы потенциалов;
3) проверку базового распределения на рациональность;
4) построение замкнутого контура и перераспределение вычислительного ресурса по контуру с целью минимизации целевой функции $F^{(\gamma)}$.

1. Построение базового распределения вычислительного ресурса. Базовое распределение вычислительного ресурса определяется матрицей $M_\varphi^{(\gamma)}$, заполненной по правилу минимальных штрафов [4], элемент $m_{\varphi_{b,i}}$ которой определяет вычислительный ресурс, выделенный заявке z_b в i-й квант времени. Для построения базового распределения вычислительного ресурса формируется матрица штрафов S_z, элемент $s_{z_{b,i}}$ которой определяет штраф при выделении заявке z_b в i-й квант времени единицы вычислительного ресурса.

Суть правила минимальных штрафов заключается в том, что заполнение матрицы $M_\varphi^{(\gamma)}$ начинается последовательно с элемента $m_{\varphi_{b,i}}$, которому в матрице штрафов S_z соответствует наименьшее значение штрафа $s_{z_{b,i}}$. Элементу $m_{\varphi_{b,i}}$ присваивается значение $\min\left(\varphi_{z_b}, \varphi_{y_i}\right)$. Изменяются значения требуемого вычислительного ресурса для обработки заявки z_b и доступного суммарного вычислительного ресурса ОВС, соответствующего i-му кванту времени: $\varphi_{z_b} = \varphi_{z_b} - \min\left(\varphi_{z_b}, \varphi_{y_i}\right)$; $\varphi_{y_i} = \varphi_{y_i} - \min\left(\varphi_{z_b}, \varphi_{y_i}\right)$.

Из дальнейшего рассмотрения в матрице S_z исключается либо строка, соответствующая заявке z_b, потребность в вычислительном ресурсе которой полностью удовлетворена, либо столбец, соответствующий i-му кванту времени, доступный суммарный вычислительный ресурс которого полностью израсходован, либо столбец и строка, если полностью израсходован доступный суммарный вычислительный ресурс для i-го кванта времени и удовлетворена потребность в вычислительном ресурсе 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 вычислительном ресурсе. Незаполненным элементам матрицы $M_{\varphi}^{(\gamma)}$ присваиваются нулевые значения. На этом построение базового распределения заканчивается.

2. Построение системы потенциалов. 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 $M_{\varphi}^{(\gamma)}$, отличных от нуля, равно $h_z + h_t - 1$ [5]. При построении базового распределения может оказаться, что ненулевых элементов матрицы $M_{\varphi}^{(\gamma)}$ меньше, чем $h_z + h_t - 1$, т.е. план распределения γ является вырожденным. В этом случае нулевым элементам матрицы $M_{\varphi}^{(\gamma)}$ с наименьшим значением $s_{z_{b,i}}$ последовательно присваиваются значения условного базисного нуля до тех пор, пока число элементов матрицы $M_{\varphi}^{(\gamma)}$, отличных от нуля, не станет равно $h_z + h_t - 1$.

Каждому i-му кванту времени ставится в соответствие потенциал p_{t_i}, а каждой заявке z_b – потенциал p_{z_b}. Потенциалы p_{t_i} и p_{z_b} выбираются таким образом, чтобы для каждого ненулевого элемента матрицы $M_{\varphi}^{(\gamma)}$ выполнялось условие $p_{z_b} + p_{t_i} = s_{z_{b,i}}$, а для нулевого элемента матрицы $M_{\varphi}^{(\gamma)}$ – условие $p_{z_b} + p_{t_i} \le s_{z_{b,i}}$. Так как число всех потенциалов равно $h_z + h_t$, а отличных от нуля элементов матрицы $M_{\varphi}^{(\gamma)}$ – $h_z + h_t - 1$, то для определения значений p_{z_b} и p_{t_i} решается система уравнений $p_{z_b} + p_{t_i} = s_{z_{b,i}}$ с $h_z + h_t$ неизвестными, для чего одному из неизвестных присваивается произвольное значение (тогда система имеет единственное решение).

3. Проверка базового распределения на рациональность. Для каждого из нулевых элементов матрицы $M_{\varphi}^{(\gamma)}$ вычисляется значение его базисной оценки, определяемой разностью между величиной штрафа $s_{z_{b,i}}$ и суммой потенциалов $p_{z_b} + p_{t_i}$, соответствующих данному элементу. План является рациональным в том случае, если полученные значения базисных оценок $r_{s_{b,i}}$ для всех нулевых элементов $m_{\varphi_{b,i}}$ матрицы $M_{\varphi}^{(\gamma)}$ будут неотрицательными, т.е. $r_{s_{b,i}} = s_{z_{b,i}} - \left(p_{z_b} + p_{t_i}\right) \geq 0$.

Если хотя бы одна из вычисленных базисных оценок имеет отрицательное значение, осуществляется текущее перераспределение вычислительного ресурса. Для этого, из матрицы $M_{\varphi}^{(\gamma)}$ выбирается элемент $m_{\varphi_{b,i}}$, для которого базисная оценка $r_{s_{b,i}}$ принимает минимальное отрицательное значение. Для выбранного элемента $m_{\varphi_{b,i}}$ в матрице $M_{\varphi}^{(\gamma)}$ строится замкнутый контур. Замкнутый контур представляет собой последовательность элементов матрицы $M_{\varphi}^{(\gamma)}$, 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.

4. Построение замкнутого контура и перераспределение вычислительного ресурса. Нулевой элемент $m_{\varphi_{b,i}}$ матрицы $M_{\varphi}^{(\gamma)}$, для которого $r_{s_{b,i}}$ принимает минимальное отрицательное значение, выбирается первым элементом замкнутого контура. Остальные элементы замкнутого контура выбираются из элементов матрицы $M_{\varphi}^{(\gamma)}$, отличных от нуля.

Выбрав в качестве исходной одну из координат первого элемента контура, например, координату столбца матрицы $M_{\varphi}^{(\gamma)}$, анализируются строки матрицы в этом столбце. Следующий, отличный от нуля элемент матрицы $M_{\varphi}^{(\gamma)}$ 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 ненулевой элемент, то второй найденный элемент удаляется из контура и в столбце матрицы $M_{\varphi}^{(\gamma)}$, заданном первым элементом контура, ищется другой ненулевой элемент, который и будет вторым элементом контура. Координата строки найденного элемента указывает строку матрицы $M_{\varphi}^{(\gamma)}$, 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 $m_{\varphi_{b,i}}$, которому соответствует наименьшее значение. Значение элемента $m_{\varphi_{b,i}}$ прибавляется к значениям элементов замкнутого контура с положительным знаком и вычитается из значений элементов замкнутого контура с отрицательным знаком. В результате получим новое распределение вычислительного ресурса.

Для нового распределения вычислительного ресурса вновь строится система потенциалов и вычисляются значения базисных оценок для нулевых элементов матрицы $M_{\varphi}^{(\gamma)}$. В случае отсутствия отрицательных значений базисных оценок для нулевых элементов матрицы $M_{\varphi}^{(\gamma)}$ в новом распределении вычислительных ресурсов, принимается решение о том, что полученное распределение рационально.

Выводы. Таким образом, основным полученным научным и практическим результатом данного исследования является разработанная математическая модель, позволяющая построить равномерное распределение выделенного вычислительного ресурса для обработки множества заявок абонентов ОВС по квантам заданного интервала времени.

ЛИТЕРАТУРА

1. *Максименков А.В. Основы проектирования информационно-вычислительных систем и сетей ЭВМ. – М.: Радио и связь, 1991. – 319 с.*
2. *Королёв А.В., Кучук Г.А., Пашнев А.А. Управление сетевыми ресурсами. – Х.: ХВУ, 2004. – 272 с.*
3. *Королёв А.В., Кучук Г.А., Пашнев А.А. Адаптивная маршрутизация в корпоративных сетях. – Х.: ХВУ, 2004. – 272 с.*
4. *Майника Э. Алгоритмы оптимизации на сетях и графах: Пер. с англ. / Под ред. Е.К. Масловского. – М.: Мир, 1981. – 321 с.*
5. *Таха Х. Введение в исследование операций. – М.: Мир, 1985. – Ч. 1. – 479 с.*

Поступила 28.04.2004

***ПАШНЕВ Андрей Анатольевич**, канд. техн. наук, научн. сотр. ИВЦ ХВУ. В 1993 году окончил Харьковское высшее военное авиационное училище радиоэлектроники. Об-*

ласть научных интересов – системы обработки и передачи данных.

КУЧУК Георгий Анатольевич*, канд. техн. наук, ст. научн. сотр., начальник НИО ИВЦ ХВУ. Окончил мехмат Московского государственного университета в 1977 году. Область научных интересов – обработка информации.*

ЛЕБЕДЕВА Ирина Анатольевна*, с.н.с. ИВЦ ХВУ. Окончила факультет прикладной математики ХИРЭ. Область научных интересов – современный анализ.*