

УДК681.324

ВАРИАЦИОННЫЙ АНАЛИЗ ОРГАНИЗАЦИИ ВЫЧИСЛИТЕЛЬНОЙ СЕТИ

Скатков А.В., Нюнькина Ю.П.

Приводится подход к выбору оптимальной в смысле некоторого критерия вычислительной сети, позволяющий значительно экономить время при анализе архитектуры вычислительных сетей и не требующего разработки специальных имитационных моделей.

Выбор оптимальной структуры вычислительной сети в общем случае не может быть отнесен к классу формально разрешимых задач и составляет главное содержание творческой деятельности инженера-проектировщика.

Рассмотрим следующую задачу. Необходимо синтезировать структуру вычислительной сети с отказами при условии, что отказ в обслуживании получит в среднем не более η % заявок. В систему поступает простейший поток заявок, время обслуживания распределено по показательному закону. Задана производительность процессоров, равная B вычислительных операций в единицу времени. Процессоры считаются однородными. Для входного потока заданы интенсивность λ заявок в единицу времени и трудоемкость θ вычислительных операций.

Далее для представления модели вычислительной сети используется терминология систем массового обслуживания (СМО).

Требуется определить минимально необходимое количество процессоров n и структуру их объединения в вычислительную сеть. Рассматриваются три возможных варианта построения сети.

Первый вариант – многоканальная СМО с числом каналов n . Всякая заявка, поступившая в систему в некоторый момент времени, либо начинает сразу обслуживаться, либо теряется, если в момент ее поступления все процессоры заняты. Для многоканальной СМО вероятность отказа в обслуживании поступившего требования [1]:

$$P_n = \frac{\rho^n}{n!} / \sum_{k=0}^n \frac{\rho^k}{k!}, \quad (1) \quad (1)$$

где $\rho = \lambda / \mu$.

Это вероятность того, что все n каналов заняты.

Второй вариант – использование последовательно расположенных групп процессоров. В такой системе заявки, которые не были сразу обслужены процессорами первой группы, поступают на процессоры второй группы. Заявки, которые не были обслужены и процессорами второй группы, поступают на процессоры третьей группы и т. д. Заявка обслуживается группой процессоров, если в группе имеется хотя бы один свободный про-

цессор. Обслуженная заявка покидает систему. Пусть имеется i групп, а процессоров в каждой группе r_j , $1 \leq j \leq i$, причем $\sum_{j=1}^i r_j = n$. Для такой системы вероятность отказа в обслуживании поступившего требования [1]:

$$p_i = \frac{\rho^{r_1+r_2+\dots+r_i}}{(r_1+r_2+\dots+r_i)!} / \sum_{k=0}^{r_1+r_2+\dots+r_i} \frac{\rho^k}{k!}. \quad (2) \quad (2)$$

Это вероятность прохода необслуженной заявки сквозь всю систему, состоящую из i групп. С учетом того, что $\sum_{j=1}^i r_j = n$, вероятности отказа в обслуживании системами первого и второго типа равны.

Третий вариант организации вычислительной сети – одноканальная многофазная (n -фазная) система с отказами. Задача (заявка) разбивается на подзадачи меньшей трудоемкости, и эти подзадачи последовательно решаются на последовательных процессорах. Если заявка, поступившая на некоторую фазу обслуживания, находит соответствующий процессор занятым, то она выходит из сферы обслуживания системы и теряется. Применяя результаты, приведенные в [1], можно показать, что для такой системы вероятность отказа в обслуживании поступившего требования:

$$p_n = 1 - \mu^n [(2\mu)^{n^2-n} + (\lambda + \mu)^{n-2} \lambda^{n-2} (\mu)^{n-2}] / (n(\lambda + \mu)(2\mu)^{n^2-n}). \quad (3) \quad (3)$$

Для многоканальной системы и системы с последовательно расположенными группами процессоров вероятность обслуживания заявки одинакова, не зависит от структуры и определяется только количеством используемых процессоров. Различия между этими двумя структурами систем заключаются в интенсивности использования различных групп процессоров.

Для многоканальной системы вероятность того, что ни один процессор не занят обслуживанием, $p_0 = 1 / \sum_{k=0}^n \frac{\rho^k}{k!}$. Среднее количество занятых

процессоров $N_3 = \sum_{k=1}^n k p_k = \sum_{k=1}^n \frac{\rho^k}{(k-1)!} p_0$. Коэффициент занятости процессоров (отношение среднего числа занятых процессоров к их общему числу) $K_3 = N_3 / n$.

Для системы последовательно расположенных групп процессоров вероятность занятости процессоров j -й группы

$p_j = \frac{\rho^{r_1+r_2+\dots+r_j}}{(r_1+r_2+\dots+r_j)!} / \sum_{k=0}^{r_1+r_2+\dots+r_j} \frac{\rho^k}{k!}$. Вероятность того, что заявка будет обслужена процессорами j -й последовательной группы, равна

$P_{\text{обсл}} = (p_{j-1} - p_j) / p_{j-1}$. Коэффициент производительности процессоров j -й последовательной группы $\gamma_j = (p_{j-1} - p_j) / (1 - p_i)$. Среднее количество занятых процессоров $-N_z = \sum_{k=1}^i r_i p_k$. Коэффициент занятости процессоров (отношение среднего числа занятых процессоров к их общему числу) $K_z = N_z / n$.

Таким образом, на первом этапе решения задачи необходимо, основываясь на значениях критерия (1) - (3), сделать выбор между многофазной системой и двумя другими вариантами. Затем в случае необходимости производится выбор между многоканальной системой и системой последовательно расположенных групп процессоров с учетом коэффициента занятости процессоров.

Проиллюстрируем выбор структуры для системы со следующими параметрами. Максимальное количество необслуженных заявок $\eta = 5\%$. Трудоемкость поступающих в систему заявок $\theta = 100000$ вычислительных операций. Производительность процессоров $B = 45000$ вычислительных операций в единицу времени. Требуется синтезировать структуру вычислительной сети для интенсивности входного потока $\lambda = 0,35$ заявок в единицу времени и выяснить, какие изменения должны быть внесены в структуру при изменении интенсивности входного потока в диапазоне $0,05 - 0,50$ заявок в единицу времени.

Интенсивность обслуживания $\mu_{1,2}$ для многоканальной системы и системы последовательно расположенных групп процессоров определяется отношением производительности процессора и трудоемкостью заявки:

$$\mu_{1,2} = B / \theta = 0,45.$$

Для многофазной системы, состоящей из n процессоров при условии что заявка разбита на n подзадач равной трудоемкости, интенсивность обслуживания определяется соотношением:

$$\mu_3 = \frac{B}{(\theta / n)} = \frac{B \cdot n}{\theta} = \mu_{1,2} \cdot n = 0,45n.$$

Основными критериями для выбора структуры вычислительной сети в данной задаче являются ограничение на вероятность обслуживания заявки (минимум 0,95) и минимум используемых процессоров.

Результат численного анализа зависимости количества обслуженных заявок от интенсивности входного потока представлен на рисунках 1 и 2. Значения интенсивности потока изменяются в заданном диапазоне 0,05... 0,50. Графики иллюстрируют системы, состоящие из одного, двух, трех и четырех процессоров. На рисунке 1 изображены зависимости для много-

фазной системы, на рисунке 2 – для многоканальной системы и системы последовательно расположенных групп процессоров.

Из рисунков видно, что при интенсивности входного потока в диапазоне 0,5...0,15 при любой организации вычислительной системы требуется два процессора для выполнения требования к вероятности обслуживания заявки. Предпочтительно не использовать в данном случае

двухфазную систему, несмотря на то, что вероятность отказа такой системы в рассматриваемом диапазоне интенсивностей меньше, так как для ее реализации необходимо дополнительно разбивать заявку на подзадачи.

При интенсивности входного потока в диапазоне 0,15...0,25 для реализации многофазной системы при условии выполнения ограничения на вероятность отказа требуется два процессора, в то время как для реализации других возможных структур систем необходимы три процессора.

При интенсивности входного потока в диапазоне 0,25...0,40 для выполнения ограничения на вероятность отказа необходимо реализовывать многоканальную систему или систему последовательно расположенных групп процессоров, состоящую из трех процессоров, а при интенсивности 0,40...0,50 – из четырех процессоров.

Таким образом, для входного потока интенсивностью 0,35 заявок в единицу времени, для выполнения ограничения на вероятность отказа, необходимы три процессора. В этом случае вероятность отказа в обслуживании заявки составляет 0,03. Процессоры могут объединяться в многоканальную систему или систему последовательно расположенных групп процессоров. Рассмотрим

$P_{..}$

1 – один процессор 2 – два процессора
3 – три процессора 4 – четыре процессора
Рисунок 1 – Зависимость вероятности обслуживания заявки многофазной системой от интенсивности входного потока заявок λ

$P_{..}$

1 – один процессор 2 – два процессора
3 – три процессора 4 – четыре процессора
Рисунок 2 – Зависимость вероятности обслуживания многоканальной системой и системой последовательно расположенных групп процессоров от интенсивности входного потока заявок λ

более подробно параметры этих двух видов систем, определяющие эффективность использования процессоров.

Для многоканальной системы вероятность того, что ни один процессор не занят обслуживанием, равна 0,46. Коэффициент занятости процессоров (отношение среднего числа занятых процессоров к их общему числу) равен 0,25. Среднее количество занятых процессоров 0,75.

Для системы последовательно расположенных групп процессоров, состоящей из трех процессоров, возможны три следующих варианта структуры: два процессора в первой группе и один во второй; один процессор в первой группе - два во второй; три последовательных группы по одному процессору в группе.

Рассмотрим структуру один процессор – два процессора. Для такой структуры вероятность обслуживания заявки сразу же, на первом процессоре равна 0,56. Коэффициент использования первого процессора – 0,58, т.е. 58% всех заявок обслуживается на первом процессоре. Среднее количество занятых процессоров 0,32. Для одного процессора среднее количество занятых процессоров означает загрузку этого процессора. Во второй группе, состоящей из двух процессоров, среднее количество занятых процессоров равно 0,31. Общее среднее количество занятых процессоров в двух группах – 0,63.

Для структуры «два процессора – один процессор» значения параметров следующие: вероятность обслуживания заявки сразу же, на одном из двух процессоров первой группы равна 0,85. Коэффициент использования двух процессоров первой группы 0,89, т.е. эти процессоры обслуживают 89% заявок. Среднее количество занятых процессоров в этой группе равно 0,60. Во второй группе, состоящей из одного процессора, среднее количество занятых процессоров – 0,08. Общее среднее количество занятых процессоров в двух группах 0,68.

Рассмотрим также систему, состоящую из трех групп по одному процессору в каждой группе. Вероятность обслуживания заявки на первом же процессоре 0,56. На первом процессоре обслуживается 58% всех заявок. Среднее количество занятых процессоров 0,32. Значения параметров совпадают со значениями параметров для первого процессора в системе со структурой «один процессор – два процессора». Второй процессор, составляющий вторую группу, обслуживает 30% всех заявок. Для него среднее количество занятых процессоров – 0,19. Для последнего, третьего, процессора среднее количество занятых процессоров – 0,08. Суммарное среднее количество занятых процессоров для системы данной структуры – 0,59.

Таким образом, достижение высокой эффективности вычислительной системы достигается за счет структурной избыточности, т.е. неэффективного использования процессоров. Наивысший показатель загрузки процессо-

ров имеет многоканальная система, причем в такой вычислительной системе нагрузка распределена между процессорами равномерно.

В системе с последовательно расположенными группами процессоров эффективность использования процессоров меньше, чем в многоканальной системе, и разная для различных структур групп. В порядке убывания эффективности использования процессоров следуют системы: два процессора – один процессор; один процессор – два процессора; три группы по одному процессору.

В системе с последовательными группами процессоров вычислительная нагрузка между процессорами распределена неравномерно. Системы такой структуры могут быть предпочтительнее, например, в случае если процессоры имеют неодинаковые показатели надежности.

В данной задаче процессоры считаем однородными. Таким образом, с использованием вариационного анализа для заданных численных значений параметров выбран оптимальный вариант структуры – многоканальная вычислительная система.

Использование предлагаемого подхода позволяет значительно экономить время при предпроектном анализе архитектуры вычислительных сетей и не требует разработки специальных имитационных моделей. Предлагаемый подход может также быть применен, например, при решении задач выбора структуры технологических линий автоматизированного производства.

Библиография

1 Новиков О.А. Прикладные вопросы теории массового обслуживания / О.А. Новиков, С.И. Петухов. — М.: Сов. радио, 1969. — 400 с.

Поступила в редакцию 02.11.2000 г.